'VOIR' LA PAROLE

Michel Chafcouloff

Résumé

Dans la présente étude sont décrits les différents procédés élaborés par les hommes de science et les chercheurs pour donner une représentation graphique de la parole dans le domaine de l'acoustique. Des méthodes mécaniques les plus anciennes jusqu'aux méthodes les plus modernes fondées sur le traitement numérique du signal, nous retraçons selon un ordre chronologique l'avancée des connaissances auxquelles elles ont donné lieu dans des domaines aussi divers que ceux de la recherche phonétique et linguistique, de la recherche appliquée (synthèse et reconnaissance automatique de la parole) et de la thérapie de la parole.

Mots-clés : acoustique, analyse, kymographie, oscillographie, reconnaissance automatique, spectrographie, synthèse, thérapie.

Abstract

The present study is aimed at describing the various devices and methods worked out by scientists to give a graphic display of speech in the acoustical domain. From the most ancient methods to the most recent ones based on digital speech processing, we follow by steps the regular advance of knowledge in the various domains of phonetic and linguistic research, applied research and speech therapy.

Keywords : acoustics, analysis, kymography, oscillography, automatic speech recognition, spectrography, synthesis, speech therapy.

CHAFCOULOFF, Michel (2004), 'Voir' la parole, *Travaux Interdisciplinaires du Laboratoire Parole et Langage,* vol. 23, p. 23-65.

Introduction

De tous les systèmes de communication, la parole est pour l'être humain le moyen le plus naturel et le plus direct pour établir un contact avec son semblable, et lui faire part d'une opinion, d'une information ou d'une émotion. Toutefois l'avantage principal du message oral qui réside dans l'instantanéité de son émission et de sa transmission, a pour contrepartie son immatérialité, laquelle a longtemps constitué un obstacle rédhibitoire pour pénétrer les arcanes de la parole. Ne laissant derrière elle aucune trace de son éphémère passage dans le temps et dans l'espace, la parole a été longtemps synonyme d'abstraction et de virtualité. Dans ces conditions, comment accéder à la connaissance de ce qui n'existe que dans la mémoire ou dans la conscience humaine ? Et effectivement, si l'on ne retient que cet aspect de la matérialité, les différences qui existent entre l'écrit et l'oral expliquent les raisons pour lesquelles le premier a été longtemps privilégié par rapport au second.

Le message écrit dont la perception fait appel au sens de la vue, repose sur l'existence d'un support matériel. Que ce soit sous la forme de signes ou de symboles gravés dans la pierre au temps de la préhistoire, de tablettes d'argile sous l'antiquité, de textes manuscrits au Moyen Age ou de pages imprimées à partir de la Renaissance, le message écrit n'a pu se conserver, s'étudier et *a fortiori* se transmettre qu'en raison de sa matérialité.

Tel n'est pas le cas du message oral perçu grâce au seul sens de l'ouïe, et dont la perception par le biais de l'oreille ne nécessite pas l'existence d'un support matériel quelconque. Invisible, impalpable, immatérielle, la parole a été de tout temps opposée à l'écriture comme le rappelle de façon métaphorique le vieil adage latin '*Verba volant, scripta manent'*. Et si le propre de la parole est de s'envoler, il est une évidence qu'elle n'a pu ni se conserver sous une forme quelconque, ni être l'objet d'une étude scientifique pendant des siècles. En réalité, ce long cheminement vers la connaissance s'est fait à un rythme inégal. Sur le plan physiologique, le fonctionnement du mécanisme de production des sons a été découvert dès la plus haute antiquité par les philosophes grecs Hippocrate et Aristote, et par certains grammairiens indiens comme en témoigne un classement articulatoire des sons de la langue hindi établi par Panini dès le IVe siècle avant J.-C. Même si les premières descriptions de l'appareil phonatoire étaient incomplètes, comme en attestent les coupes schématiques du larynx et du conduit vocal esquissées par Léonard de Vinci, il est inéluctable que les fondements de la production de la parole avaient été posés de façon à la fois précoce et précise.

Sur le plan acoustique (ce terme n'a été introduit qu'au XVIIe siècle par le physicien français J. Sauveur), on doit se rendre à l'évidence que, en tant que phénomène sonore, la parole n'a pas été

décrite avant l'ère moderne. L'une des rares images où la parole est représentée sous une forme graphique, se rapporte à une fresque datant du IIe siècle avant J.-C., qui illustre un dialogue entre deux indigènes de la région de Palau en Micronésie (figure 1).

Figure 1
L'acte phonatoire symbolisé par des oscillations issues de la bouche des locuteurs.
From the Palau civilization, extrait de G. Panconcelli-Calzia (1957)

La figure ci-dessus montre que le flux d'air phonatoire issu de la bouche des deux personnages est représenté sous la forme d'un mouvement ondulatoire qui se propage dans l'air, et que l'auteur de l'article, duquel est extrait cette illustration, a assimilé aux oscillations de l'onde glottique. Toutefois, si cette peinture murale est la preuve que, même à une époque reculée, les anciens étaient capables de représenter la parole sous une forme concrète plus ou moins proche de la réalité physique, il n'en demeure pas moins que la parole est restée pendant des siècles un phénomène auréolé de mystère, et un objet d'étude inaccessible en raison de l'absence de moyens d'investigation.

Cette période d'obscurantisme scientifique dura jusqu'au XIXe siècle, période à partir de laquelle commença à se manifester un regain d'intérêt pour l'étude des langues. En effet, c'est à cette époque qu'on allait assister à la naissance d'une discipline nouvelle, la 'phonétique historique' dont l'objectif était la découverte des lois responsables de la mutation et de la transformation des sons du langage, discipline qui allait donner lieu à l'entreprise de travaux de grammaire comparée entre les langues indo-européennes.

Cependant, malgré des progrès sensibles dans la compréhension du processus d'évolution des langues, la recherche linguistique restait limitée parce qu'elle n'avait pas encore été dotée des

moyens qui allaient lui permettre de passer de l'impression subjective à l'expérience objective. Le pas décisif en la matière allait être franchi avec l'avènement de nouvelles méthodes, et en particulier celle de la méthode graphique, qui allait être introduite dans les sciences naturelles par les physiologistes et les physiciens de l'école allemande très influente en Europe durant cette deuxième moitié de siècle. L'étude de la langue qui avait été jusqu'alors l'œuvre plus ou moins exclusive des grammairiens et des philologues, allait être complétée par l'étude de la parole grâce aux phonéticiens dits 'expérimentalistes', adeptes de ladite méthode graphique qui allait leur permettre de 'voir' la parole. Et effectivement, c'est à partir du moment où le sens de la vue s'est substitué à celui de l'ouïe, c'est-à-dire à partir du moment où la parole n'a plus été seulement un phénomène 'audible', mais également 'visible' en termes de tracés représentatifs de paramètres articulatoires, aérodynamiques ou acoustiques, que la recherche phonétique a pu prendre son véritable essor.

Compte tenu de la spécificité de nos travaux antérieurs, dont la plupart se rapportent à l'acoustique des sons et la recherche de leurs principaux indices, le présent travail sera circonscrit à la présentation des méthodes de visualisation qui ont permis de décrire le signal de parole en termes de ses paramètres acoustiques. C'est ainsi que, selon un ordre chronologique, nous exposerons en un premier temps les principes de la méthode kymographique qui, bien qu'elle ait été surtout utilisée pour recueillir des informations articulatoires et aéro-dynamiques, a également servi au recueil des premières données acoustiques concernant la structure des voyelles et des consonnes voisées. Ensuite, nous nous attacherons à décrire la méthode oscillographique qui a été la première méthode d'analyse fondée sur l'utilisation du courant électrique. Enfin, nous exposerons les principes de la méthode spectrographique adoptée dans tous les centres de recherche entre 1950 et 1970, et dont l'apport a été déterminant pour la connaissance des propriétés acoustiques des sons du langage. Ensuite, nous traiterons de deux de ses principales applications : en premier lieu, dans le domaine de la communication homme-machine, la synthèse et la reconnaissance automatique ; en deuxième lieu, dans le domaine de la thérapie de la parole, la rééducation des malentendants et l'aide au diagnostic pour le traitement des dysfonctionnements de la parole.

Parallèlement, et dans le cadre que nous avons défini, nous citerons les travaux les plus marquants effectués par les chercheurs au moyen de ces appareils de visualisation du signal de parole. Nous verrons comment, depuis les systèmes mécaniques les plus anciens jusqu'aux procédures analytiques les plus modernes, chacune de ces méthodes de représentation graphique a contribué à l'avancement des connaissances dans les domaines divers, mais complémentaires de la recherche fondamentale, appliquée et thérapeutique.

1. Les méthodes de visualisation

1.1. La kymographie

Comme nous l'avons mentionné ci-dessus, c'est dans la deuxième moitié du XIXe siècle que la recherche scientifique a franchi une étape décisive grâce à l'introduction de méthodes quantitatives d'enregistrement et de mesure. Celles-ci ont été originellement appliquées à diverses disciplines des sciences naturelles, comme la météorologie, l'astronomie ou la physique. En ce qui concerne la physiologie et en particulier l'étude du fonctionnement du mécanisme de production de la parole, c'est grâce au 'Kymographon' que l'allemand Ludwig a procédé au premier enregistrement des mouvements respiratoires avec et sans parole (figure 2).

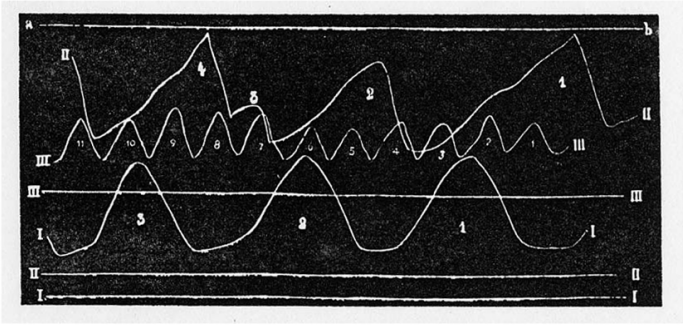


Figure 2
Premier enregistrement kymographique illustrant l'introduction de la méthode graphique en phonétique (Vierordt and Ludwig, 1855), extrait de G. Panconcelli-Calzia (1957)

Une version dérivée du 'Kymographon' a été adaptée aux besoins expérimentaux des phonéticiens, et est rapidement devenue leur outil de travail privilégié. Cet appareil qui était fondé sur un principe mécanique, était composé de deux parties :

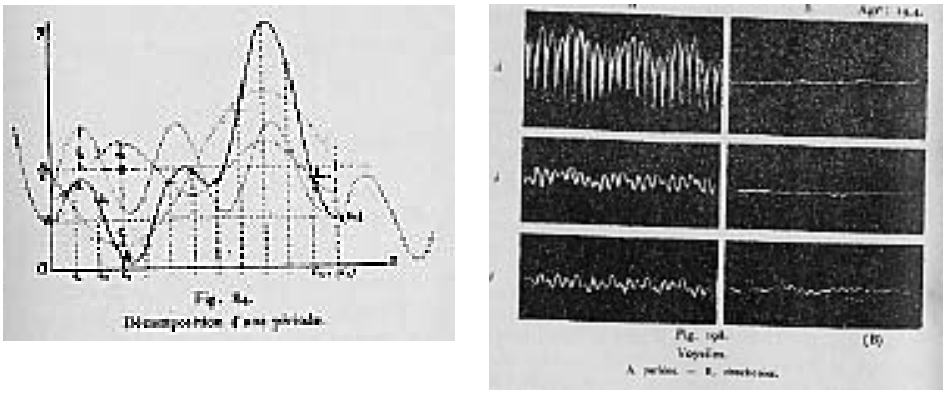
- une partie 'enregistrement' qui comprend un système pour capter le flux d'air phonatoire issu de la bouche et du nez, par l'intermédiaire d'une embouchure pour la pression buccale et d'une ou deux olives de verre pour la pression nasale. Le flux d'air est conduit grâce à des tuyaux de caoutchouc vers des membranes qui vont entrer en vibration sous l'effet de la pression d'air. Simultanément, on procède à l'enregistrement des vibrations laryngées par application d'un diaphragme sur le larynx au niveau des cordes vocales.

- une partie 'inscription' dont l'élément principal est constitué d'un cylindre (tambour de Marey) lui-même entraîné par un système de rotation mécanique. Des stylets s'appuient sur le cylindre recouvert d'un papier enduit de noir de fumée, et inscrivent des tracés dont les variations d'amplitude sont proportionnelles à la quantité d'air expiré, (pour une description de la version originale du kymographe, le lecteur se reportera à la description qui en a été faite par l'abbé Rousselot (1897), et en ce qui concerne sa version moderne, le 'polyphonomètre', (Teston, 1984)).

En fonction de sa conception, le kymographe était destiné à l'enregistrement et à l'analyse de paramètres aéro-dynamiques, en l'occurrence la pression ou débit d'air buccal, et la pression ou débit d'air nasal. Le troisième tracé qui concerne l'enregistrement des vibrations laryngées allait être utilisé pour recueillir les informations concernant la fréquence des voyelles et de certaines des consonnes voisées. À la suite des travaux du physiologiste allemand Brücke (1856), des études acoustiques allaient être entreprises sur les sons des langues les plus parlées en Europe, notamment l'anglais (Sweet, 1890), l'allemand (Viëtor, 1898), mais aussi dans des langues moins universelles comme le finno-ougrien (Pipping, 1890). En ce qui concerne le français, l'abbé Rousselot *(op. cit.)* qui est considéré comme le père fondateur de la phonétique expérimentale, a été le premier chercheur à utiliser cette méthode graphique en France.

Sur le plan physiologique, il montrera que l'articulation des consonnes comprend plusieurs phases 'constitutives' (l'attaque, la tenue et la détente) et qu'il existe des différences de force articulatoire entre les consonnes sourdes et les consonnes sonores, les premières étant articulées de façon plus énergique que les secondes. En outre, l'examen des tracés de pression orale et nasale révélera un chevauchement entre les unités phoniques sous la forme de mouvements d'anticipation et de rétroaction, première preuve concrète de la réalité du phénomène de la coarticulation dans la parole.

Sur le plan acoustique, l'abbé Rousselot montrera que les sons du langage peuvent être décrits en termes de trois paramètres, la fréquence, l'intensité et la durée. En ce qui concerne l'analyse de la fréquence, il utilisera la méthode mathématique fondée sur l'application du Théorème de Fourier. Sur la base des tracés des vibrations laryngées (figure 3), il procédera à la décomposition de l'onde périodique en ses diverses composantes pour déterminer les principales notes de résonance des voyelles orales et nasales du français. Ses calculs confirmeront le bien-fondé des hypothèses du physicien allemand Helmholtz (1863) qui avait jeté les premiers fondements de la théorie acoustique des voyelles, en démontrant que les différences de timbre étaient liées aux résonances des cavités du conduit vocal.



A *B*

Figure 3
A. Décomposition d'une période
B. Enregistrement kymographique de voyelles chuchotées et parlées
extraits de l'abbé Rousselot (1897)

En dépit d'insuffisances dont les plus manifestes étaient l'inertie du mécanisme transcripteur et les propriétés résonatrices du système, l'introduction de la kymographie allait constituer une étape décisive dans l'histoire des sciences phonétiques. En effet, c'était la première fois qu'était obtenue une représentation visuelle de la parole, grâce à laquelle on allait décrire les systèmes phonématiques des langues, non plus en termes d'impressions auditives subjectives, mais en termes de données quantitatives objectives. Cependant, les informations obtenues grâce à la méthode kymographique restaient limitées pour plusieurs raisons. Sur un plan pragmatique, la procédure d'enregistrement était longue et fastidieuse. En effet, elle comportait l'enregistrement du signal vocal avec une application plus ou moins 'hermétique' de l'embouchure, l'insertion d'une ou de deux olives nasales dans les narines, et enfin l'application du diaphragme sur le larynx du locuteur. D'autre part, la fabrication des documents nécessitait un réglage en hauteur des stylets inscripteurs, le réglage de leur force d'appui sur le rouleau, le noircissement et le vernissage final du papier etc. Enfin, inconvénient majeur, l'exactitude des mesures effectuées d'après la décomposition de l'onde périodique était sujette à caution en raison des déformations de ladite onde dues au frottement du stylet sur le rouleau. En conclusion, les contraintes techniques de l'appareillage étaient nombreuses, et l'introduction de la méthode oscillographique, première méthode de visualisation fondée sur l'utilisation du courant électrique, allait apporter des solutions nouvelles tant sur le plan de la facilité d'utilisation que sur celui de la rigueur scientifique.

1.2. L'oscillographie

L'oscillographe, premier appareil inscripteur fonctionnant à l'électricité, a fait son apparition vers les années 1920-1930. À l'origine, il n'était pas destiné à la recherche phonétique, mais grâce à la transformation de l'appareillage de base, et au couplage avec des appareils enregistreurs, il est devenu rapidement un instrument fort utile pour l'analyse des paramètres acoustiques de la parole. Par rapport au kymographe, l'oscillographe présentait plusieurs avantages sur le plan :

- de la facilité d'utilisation, car il permettait un gain de temps appréciable, du fait que le signal enregistré directement à partir d'un microphone ou à partir d'un magnétophone, était inscrit sur un rouleau de papier en sortie d'un enregistreur à jet d'encre.
- de la rigueur scientifique, il permettait d'avoir une image plus fidèle de l'onde sonore du fait que le système de reproduction n'était pas soumis aux déformations de l'onde occasionnées par le stylet inscripteur ; de plus, l'expérimentateur avait la possibilité de 'figer' le signal sur l'écran de l'oscilloscope, et de modifier sa représentation en procédant à l'expansion ou à la compression de l'onde sinusoïdale.

L'un des appareils enregistreurs les plus employés dans la recherche phonétique a été le 'Mingographe' multi-canaux, couplé à un détecteur de mélodie pour l'analyse de la fréquence fondamentale et à un intensimètre pour l'analyse de l'énergie sonore. En tant qu'appareil de visualisation, le Mingographe permettait de représenter la parole sous la forme de trois tracés relatifs à l'onde sinusoïdale, la fréquence fondamentale et l'intensité.

1.2.1. L'onde sinusoïdale

De même que Rousselot et ses collègues avaient utilisé les tracés des vibrations laryngées pour calculer les notes de résonance des voyelles, les phonéticiens des années 1930 allaient, en un premier temps, utiliser les tracés de l'onde sinusoïdale pour vérifier la précision des mesures effectuées par leurs prédécesseurs. La décomposition de l'onde périodique était faite soit d'après l'image acoustique du son à partir d'un oscilloscope à tube cathodique (figure 4), soit d'après l'image inscrite sur papier au moyen d'un inscripteur galvanométrique.

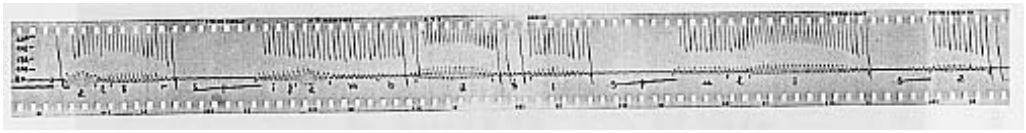


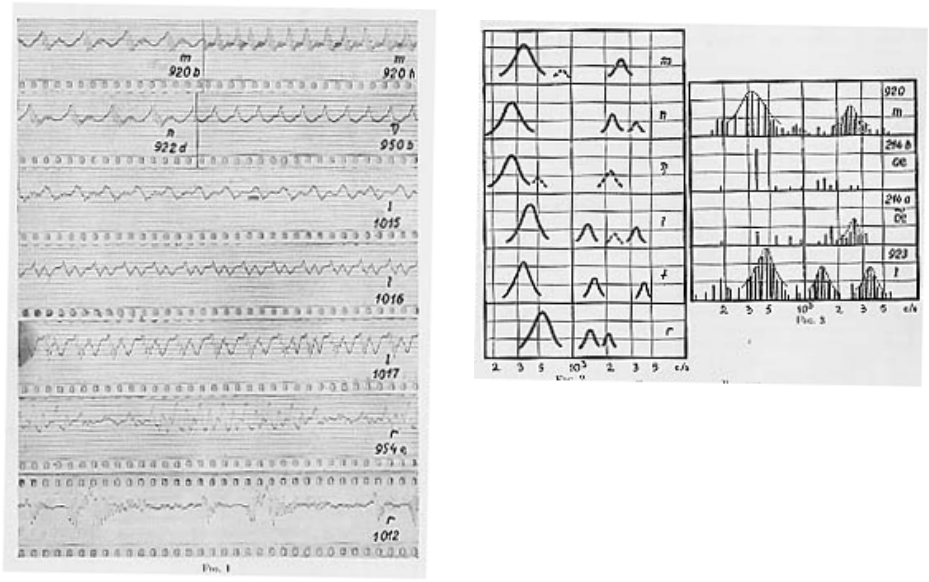
Figure 4
Oscillogramme réalisé à l'oscilloscope, extrait de Rossi (1965)

Sur le plan qualitatif, la méthode oscillographique n'a pas conduit à l'obtention de résultats radicalement différents de ceux obtenus par les utilisateurs de la méthode kymographique. De ce point de vue, on peut dire qu'elle a simplement permis de confirmer la justesse des mesures précédentes. Sur le plan quantitatif, elle allait conduire à une accélération de la recherche du fait qu'elle permettait l'analyse rapide d'un plus grand nombre de matériaux linguistiques. C'est ainsi que de nombreuses données acoustiques seront recueillies sur la fréquence fondamentale, la fréquence et la largeur de bande des formants des voyelles orales de l'anglais (Crandall, 1925), de l'allemand (Trendelenburg, 1935), du français (Grammont, 1933) et de l'italien (Gemelli & Pastori, 1934). À propos des voyelles nasales du français, on citera les travaux de Marguerite Durand (1947) alors que Merry (1921) et Sir Richard Paget (1924) trouveront des extra-résonances à 200 Hz à propos des voyelles nasalisées de l'anglais, et que Fletcher (1929) associera la nasalité à un formant bas (400 Hz) et à un formant haut situé entre 2200 et 4000 Hz.

En ce qui concerne les consonnes voisées et plus particulièrement les consonnes nasales, les chercheurs utiliseront également la méthode mathématique pour déterminer leurs caractéristiques fréquentielles. C'est ainsi que Fletcher *(op. cit.)* et Crandall *(op. cit.)* en anglais, Grammont *(op. cit.)*, en français, Sovijarvi (1938) en finlandais et Tarnoczy (1948) en hongrois, montreront que les tenues de [m, n, ɳ] sont caractérisées par la présence de pics d'énergie variables en fonction de leur lieu d'articulation. Ce dernier auteur montrera d'après l'analyse des photographies d'écran d'un oscilloscope (figure 5) que la différence acoustique majeure entre /l/ clair et /l/ sombre, réside en anglais dans la position du deuxième pic d'énergie, et que les tenues des variantes battues de /R/ sont souvent caractérisées par une alternance de segments vocaliques et de segments de bruit.

1.2.2. La fréquence fondamentale

Parmi les paramètres acoustiques qui jouent un rôle dans l'analyse des traits prosodiques, la fréquence fondamentale (F0) occupe assurément une place privilégiée. Couplé à un détecteur de mélodie, l'oscillographe à canaux a rendu possible la visualisation de la courbe de F0. De manipulation aisée, même s'il nécessitait certains réglages de calibration en fonction des différences de tessiture des locuteurs ou de la présence d'un bruit de fond pendant l'enregistrement, l'oscillographe a été durant des décennies l'instrument de travail favori des linguistes versés dans l'étude des propriétés suprasegmentales des langues. La richesse de la bibliographie exhaustive rassemblée durant les années 1970-1975 par notre collègue Di Cristo (1975) atteste de la quantité, de la qualité et de la diversité des études de prosodie, lesquelles ont porté en règle générale sur les structures intonatives des langues, les faits accentuels (accent de mot ou de phrase), le rythme, la microprosodie, la tonologie etc.



A *B*

Figure 5
A. Oscillogrammes de consonnes nasales et de consonnes vocaliques (mama, lili, ruru)
B. Courbes de résonance manuscrites
extraits de Tarnoczy (1947)

Outre le fait que le tracé de la courbe mélodique ait été utilisé pour la mesure des variations de fréquence fondamentale (figure 6), celui-ci a également servi à l'estimation de la durée dans les études, où il a été démontré que ce paramètre joue un rôle dans la perception de la proéminence accentuelle. On citera également les études qui ont porté sur la détermination de la durée des segments phoniques, des syllabes, des groupes rythmiques, des jonctures et des pauses. Enfin, on mentionnera les travaux de psycho-acoustique sur le seuil différentiel de durée (Rossi, 1972) ou encore le seuil de glissando ou seuil de perception des variations tonales (Rossi, 1971b), travaux dans lesquels les variations fines de F0 et de durée seront mesurées d'après les tracés de la courbe mélodique conjointement avec les tracés oscillographiques.

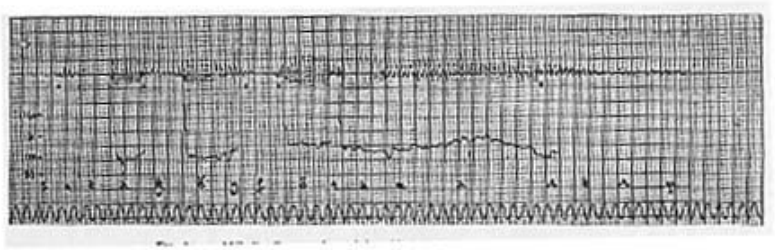


Figure 6
Oscillogramme et courbe mélodique obtenus au moyen du Mingographe, extrait de Rossi (1965)

L'avènement de dispositifs de calculs rapides par ordinateur et l'élaboration d'algorithmes fondés sur l'application du Théorème de Fourier ont été à l'origine de l'entreprise de nombreuses études destinées à améliorer l'extraction et la détection automatique de la fréquence fondamentale. Il n'est pas de notre intention, ni de notre dessein de dresser ici un inventaire de ces différentes méthodes. Parmi celles-ci, on citera brièvement la méthode du SIFT (méthode de filtrage inverse du signal), la méthode par calcul rapide du peigne spectral (Martin, 1986) ou encore la méthode de modélisation mélodique (MOMEL) fondée sur une approximation quadratique (Hirst et Espesser, 1993). Pour un inventaire complet à ce sujet, on se reportera à l'historique qui a été fait récemment par P. Martin à propos des différents systèmes élaborés par les ingénieurs pour la détection et l'analyse de la fréquence fondamentale (Martin, 2005).

Dès lors, conséquence des progrès rapides effectués dans le domaine de l'électronique, de l'informatique, ainsi que dans les méthodes de traitement numérique du signal, de nouveaux appareils de visualisation de la courbe mélodique allaient être mis à la disposition de l'enseignant (apprentissage de la prosodie), du chercheur ou du thérapeute de la parole. Parmi ces appareils, on citera plus particulièrement le *Visi-pitch* commercialisé dès 1975, mais dont la dernière version *Visi-pitch IV* est susceptible de recevoir de nombreuses applications thérapeutiques, le *Pitch-Computer* pour la visualisation en temps réel de la fréquence fondamentale et de l'intensité en 1978, le *Speech Viewer* en 1985. Parmi les logiciels, on citera plus particulièrement le logiciel d'analyse en temps réel *Win-Pitch* sous Windows (Martin, 1996) (figure 7), le logiciel *Speech Tutor* (2003) ainsi que

l'incontournable logiciel *Praat* utilisé dans la plupart des laboratoires de recherche sur les faits prosodiques des langues.

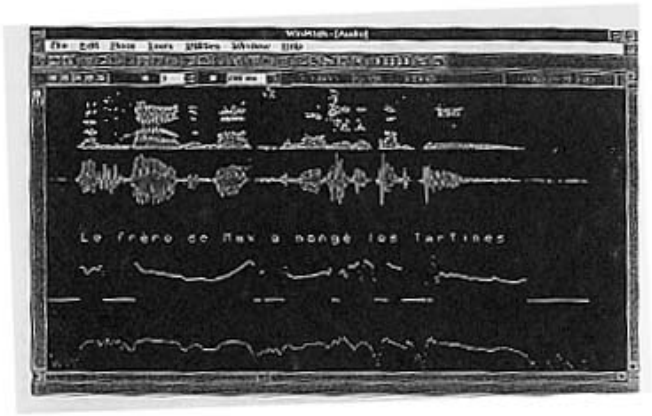


Figure 7
Visualisation de la courbe mélodique, du spectrogramme, de l'intensité et du texte, avec Winpitch sous Windows, extrait de Martin (1996)

1.2.3. L'intensité

Parmi les paramètres acoustiques qui interviennent dans la description phonétique des sons du langage, l'intensité est le paramètre qui s'est prêté le plus tardivement à l'analyse. À ce propos, il n'est pas inutile de rappeler que sur le kymographe de type classique, la ligne buccale ne se rapportait pas à l'intensité, mais exprimait seulement une valeur moyenne de la pression buccale pendant l'articulation du son. Dans ces conditions, il a fallu attendre les années 1920 et la fabrication d'appareils électriques pour voir apparaître l'intensimètre, c'est-à-dire un appareil capable de produire un voltage qui représentait, mais n'était pas nécessairement proportionnel à l'intensité de l'onde sinusoïdale. Brièvement décrit, cet appareil était composé de filtres, dont un filtre de pré-emphase, un filtre de lissage, un filtre de rectification et une unité de compression pour la représentation logarithmique de la courbe. Comme il a été établi que le niveau d'intensité d'une conversation courante était d'environ 55-60 dB, l'échelle moyenne était fixée entre 40 et 80 dB. En ce qui concerne sa représentation graphique (figure 8), l'intensité d'un segment de la chaîne parlée est exprimée en termes de surface d'aire à l'intérieur de la courbe pendant l'intervalle

temporel du segment en question. En recherche phonétique, l'intensimètre couplé à un inscripteur graphique a été utilisé dans plusieurs types d'études :

- les études sur les structures prosodiques des langues, dans lesquelles a été déterminé le rôle de l'intensité dans la perception de l'accent, son intégration temporelle, ainsi que ses relations avec les autres paramètres dans la perception des schémas intonatifs.
- les études sur la production des unités segmentales, dans lesquelles a été déterminé le niveau d'intensité spécifique des voyelles (Rossi, 1971a), ainsi que les différences d'intensité entre les voyelles et les consonnes dans une optique d'application à la synthèse par règles et à la reconnaissance automatique.
- les études de psycho-acoustique où l'analyse des variations fines de ce paramètre a été effectuée d'après les tracés d'intensité globale, et plus particulièrement dans les travaux sur la perception de la sonie et la détermination du seuil différentiel d'intensité des voyelles (Rossi, 1976b, 1978).

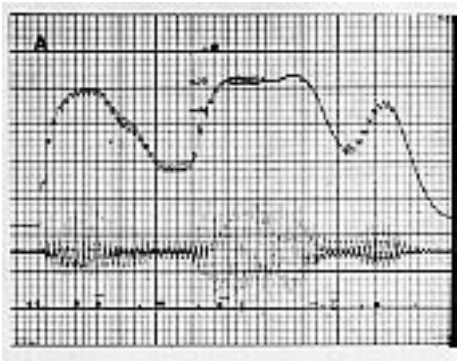


Figure 8
Oscillogramme et courbe d'intensité obtenus au moyen du Mingographe, extrait de Rossi (1976a)

Parallèlement à la courbe mélodique, la visualisation de la courbe d'intensité (synonyme de puissance ou d'énergie par unité de temps) a été facilitée par l'avancement technologique et informatique, et la plupart, si ce n'est tous les logiciels élaborés à cette époque, ont été également conçus pour la détection et la visualisation de la courbe d'intensité. À partir de cet instant, et contrairement à leurs prédécesseurs, les chercheurs des années du 'boom' informatique allaient pouvoir bénéficier de conditions privilégiées, et englober d'un coup d'œil, l'ensemble des paramètres acoustiques qui jouent un rôle important dans la production et la perception de la parole.

1.3. La spectrographie

À vrai dire, les tracés obtenus jusqu'alors par la méthode oscillographique ne permettaient que d'obtenir une représentation bi-dimensionnelle 'amplitude-temps' de la parole, c'est-à-dire une image des variations de pression sonore de l'onde en fonction de la durée. De plus, cette représentation était inadaptée pour établir une distinction entre les voyelles, du fait que les différences de formes des vibrations glottales étaient souvent trop fines pour être discernées même par un œil exercé. En outre, l'appareillage était inapproprié pour représenter les sons sous leur forme spectrale, c'est-à-dire pour fournir une information sur la distribution de l'énergie en fonction de la fréquence. Afin de pallier cette insuffisance, les chercheurs ont eu recours à une procédure manuelle pour donner une image des formes acoustiques des sons. La figure 9 illustre la procédure par laquelle la forme spectrale des voyelles a été reconstituée en termes de leurs trois premières résonances grâce à une succession d'analyses effectuées en différents points du continuum sonore. Si on ajoute la troisième dimension d'intensité (elle-même proportionnelle à la largeur et à la noirceur relative des concentrations d'énergie), on obtient une représentation tri-dimensionnelle de la parole en termes des trois paramètres intensité, fréquence, durée qui préfigure la méthode spectrographique qui sera introduite dès le début des années 1940.

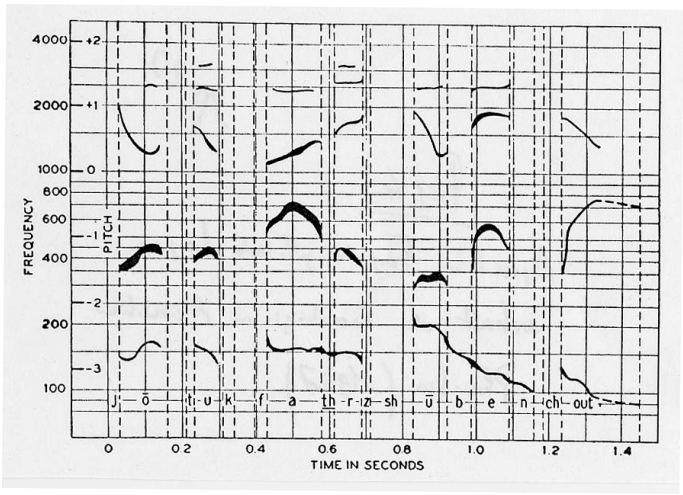


Figure 9
Variation des résonances vocales à partir de l'analyse des harmoniques de différentes périodes en fonction du temps, extrait de Steinberg (1934) dans Koenig et al. *(1946)*

La difficulté à lire un oscillogramme était due au fait que l'information concernant un son de parole était trop condensée pour en permettre l'identification. Pour que cette information soit interprétable par l'œil, il fallait concevoir une procédure capable d'effectuer une opération semblable à celle effectuée par l'oreille, c'est-à-dire d'étaler les dimensions de la parole dans le temps. Cette procédure, connue sous le nom de spectrographie, allait consister à visualiser les formes acoustiques des sons du langage, ce qui équivalait en quelque sorte à effectuer une traduction visuelle de la parole, d'où le terme de '*Visible Speech'* habituellement utilisé en langue anglaise. Quoiqu'elle ait eu à l'origine un objectif militaire bien précis, qui était de transmettre un message sous une forme visuelle et non plus orale, la spectrographie a été surtout connue pour sa fonction d'analyse du signal de parole. Le premier appareil, le '*Sonagraph*', était un analyseur de fréquence à fonctionnement successif, c'est-à-dire capable d'analyser successivement les composantes individuelles d'une onde par variation de la fréquence d'analyse d'un filtre. Le signal sonore était enregistré sur un disque magnétique, dont la rotation était synchronisée avec un cylindre recouvert d'une feuille de papier conductrice d'électricité. Un stylet inscrivait sur le papier des traces d'opacité variable qui étaient la représentation graphique de l'analyse fréquentielle des oscillations simples de l'onde complexe. La figure 10 illustre les deux principales représentations graphiques obtenues soit en filtre large à 300 Hz, soit en filtre étroit à 45 Hz.

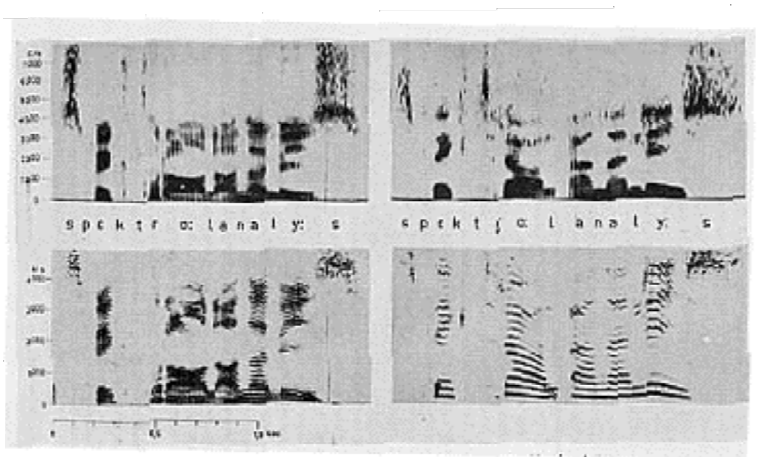


Figure 10
Spectrogrammes d'une voix masculine (à gauche) et d'une voix féminine (à droite) réalisés en filtre large (en haut) et en filtre étroit (en bas), extrait de Fant (1968)

Connue de tous les chercheurs grâce à l'ouvrage de Potter *et al.* (1947), la méthode spectrographique analogique allait être utilisée intensivement dans tous les centres de recherche sur la parole pendant les deux décennies de 1950 à 1970. C'est à partir de cette dernière date que l'on assiste, parallèlement avec l'avènement de l'informatique et le développement des méthodes de traitement numérique du signal, au retrait et plus tard à la disparition des spectrographes analogiques, qui allaient progressivement céder la place aux spectrographes numériques. La compagnie KAY Elemetrics, dont le nom est étroitement associé à tout ce qui concerne la visualisation des formes spectrales du signal de parole, allait commercialiser plusieurs systèmes comprenant différents logiciels, parmi lesquels les systèmes CSL (Computerized Speech Lab) et MS (Multi-Speech).

Les spectrographes numériques sont de véritables stations de travail pour l'acquisition, l'affichage et le traitement du signal de parole à l'intention de l'acousticien, du phonéticien ou du thérapeute. Ils sont capables de fournir en temps réel différentes images du signal sous forme de l'onde sinusoïdale, des patrons spectraux, de la courbe mélodique et de la courbe d'amplitude RMS. En outre, sont implantées les différentes méthodes d'analyse automatique comme l'analyse par FFT (Transformée de Fourier rapide), l'analyse cepstrale ou encore l'analyse par LPC (codage par prédiction linéaire) etc.

Outre leur fonction de représentation graphique et d'analyse, ces appareils permettent :

- l'affichage des paramètres prosodiques et plus particulièrement de la courbe de F0 avec la possibilité de comparer la courbe originale et la courbe synthétisée sur une même fenêtre.
- la modification des paramètres prosodiques en intervenant soit sur la fonction graphique, soit sur les tableaux numériques.
- la modification de la vitesse du débit de parole et sa relecture immédiate.
- l'affichage de deux spectrogrammes en temps réel et en mode partagé pour comparer les caractéristiques des deux signaux, etc. En fait, la potentialité de l'outil est tellement grande que ses possibilités ne sont limitées que par l'imagination de l'opérateur.

En plus de leurs fonctions d'analyse acoustique, les spectrographes numériques peuvent être utilisés pour l'extraction et l'analyse des paramètres physiologiques. C'est ainsi qu'ils peuvent être couplés avec différents périphériques comme un laryngographe pour la visualisation du mouvement des cordes vocales, un nasomètre pour la détection et l'affichage de la pression nasale, un appareil de mesure du débit d'air pour l'évaluation de la pression intraorale et autres paramètres aéro-dynamiques, ou encore un palatographe pour la visualisation en temps réel des appui linguo-palatins en synchronisation avec l'analyse par LPC (codage par prédiction linéaire) (figure 11). Les documents peuvent être imprimés soit en noir avec différentes nuances de gris, soit en couleur.

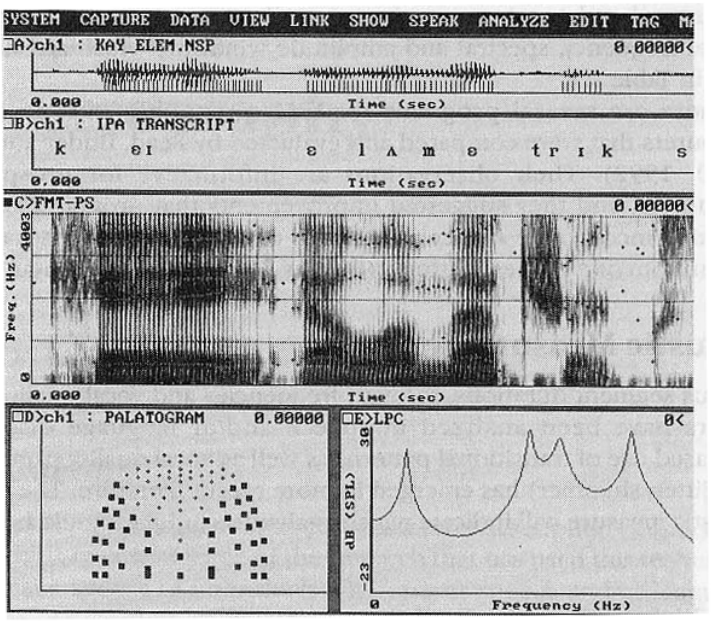


Figure 11
Visualisation d'un échantillon de parole sur spectrographe numérique KAY CSL : oscillogramme, transcription phonétique IPA, palatogramme, analyse par LPC, extrait de A. Farmer (1977)

Parallèlement à l'apparition de ces nouveaux outils, les informaticiens allaient travailler au développement de stations de travail fondées sur l'utilisation de logiciels pour l'acquisition, la numérisation, la visualisation et l'analyse du signal de parole. Là encore, ces logiciels élaborés dans la plupart des centres de recherche à travers le monde sont trop nombreux pour essayer d'en donner une liste même limitative.

Au Laboratoire Parole et Langage d'Aix-en-Provence, la version 3.2. de l'environnement logiciel SESANE (Software Environment for Speech Analysis and Evaluation) a été implantée sur les stations EVA et DIANA, qui sont des matériels d'investigation clinique pour l'aide au diagnostic et à la rééducation des dysfonctionnements de la voix et de la parole (Teston & Galindo, 1995). Les différents logiciels permettent d'appliquer des protocoles d'analyse physiologique, et de traiter des données aérodynamiques (entre autres, celles relatives au débit d'air oral/nasal (figure 12), ou à la fuite glottique) recueillies grâce aux dispositifs mentionnés ci-dessus. Les différentes mesures sont présentées sous la forme de tableaux et de diagrammes représentatifs des paramètres de la voix de sujets pathologiques par rapport à des données de sujets dits normaux.

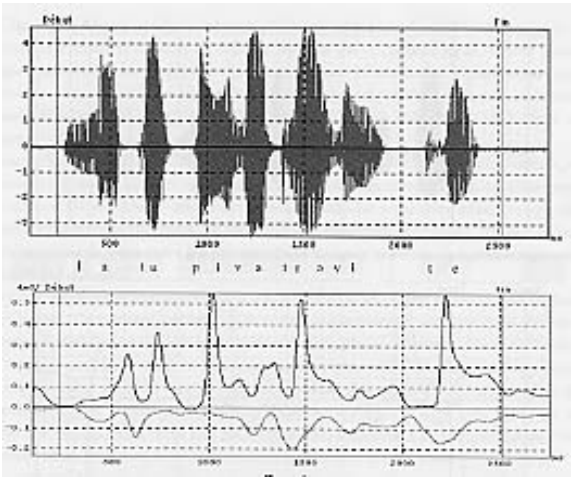


Figure 12
Signal acoustique et courbes de variation du débit d'air oral et du débit d'air nasal, extrait de B. Teston (2000a)

En ce qui concerne la spectrographie, qu'elle soit analogique ou numérique, nous distinguerons trois domaines de recherche :
- la recherche fondamentale sur la structure acoustique des sons du langage.
- la recherche appliquée aux technologies vocales, en particulier à la synthèse et à la reconnaissance automatique de la parole.
- la recherche thérapeutique pour la rééducation des malentendants et l'aide au diagnostic des troubles de la voix.

2. Les domaines de recherche

2.1. La recherche fondamentale en acoustique

C'est dans le domaine de la découverte des indices acoustiques des sons du langage que l'apport de la méthode spectrographique a été le plus spectaculaire. Pour la première fois se trouvaient rassemblées sur un même document les informations concernant les trois paramètres acoustiques de base. La durée se lisait de droite à gauche, la fréquence de bas en haut, et l'intensité était proportionnelle au degré de noirceur des zones de concentration d'énergie. En un premier temps, l'intérêt des chercheurs s'est porté sur les propriétés spectrales des voyelles.

2.1.1. Les voyelles

Selon la théorie de la résonance de Helmholtz *(op. cit.)*, le timbre spécifique des voyelles était dû à l'existence de zones d'harmoniques amplifiés appelés également formants selon la terminologie originale de Hermann (1895). D'après cette théorie, l'air expiré se mettait à vibrer dans les cavités du conduit vocal à une fréquence correspondant à la fréquence propre de chaque cavité. C'est dans la visualisation et la mesure des fréquences des formants que la spectrographie allait se révéler d'une grande utilité.

En ce qui concerne la visualisation des formants, cette méthode allait permettre de vérifier le bien-fondé de la théorie de la résonance en montrant que les formants constituaient effectivement une réalité acoustique, et apparaissaient sous la forme de barres de résonances vocales. C'est effectivement ce qu'allaient montrer *de visu* Potter *et al. (op. cit.)* en dressant l'inventaire des formes spectrales des voyelles et des diphtongues de l'anglo-américain. Selon ces auteurs, la caractéristique acoustique primaire des voyelles se situait au niveau de la deuxième barre de résonance appelée 'hub', qui constituait l'indice primaire de distinction entre les voyelles orales (figure 13).

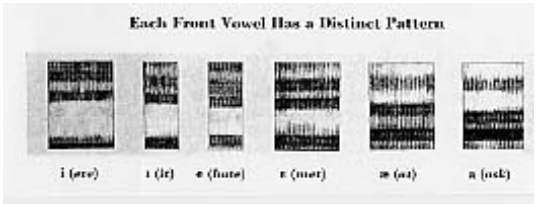


Figure 13
Spectrogramme en bande large des voyelles de l'anglo-américain illustrant la position du 'hub' définie comme la deuxième résonance vocale, extrait de Potter et al. *(1947)*

Sur la base de ces informations visuelles, et suite à une expérience de synthèse dans laquelle il était démontré que la position fréquentielle des deux premiers formants était suffisante pour caractériser chaque voyelle du point de vue de son timbre (Delattre, 1951), une représentation sur un plan F1/F2 allait être adoptée par les chercheurs, où la disposition des voyelles sur le triangle acoustique rappelle celle des voyelles sur le triangle articulatoire de la phonétique classique (figure 14).

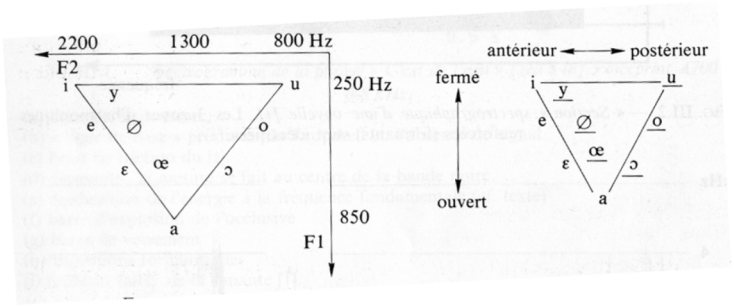


Figure 14
Relation acoustico-articulatoire des voyelles orales du français, extrait de Calliope (1989)

Compte tenu de l'ampleur des écarts de formants (en termes de valeurs en Hz) entre les voyelles, une échelle logarithmique a été couramment utilisée pour les axes F1/F2. Cependant, afin de mieux rendre compte des distances subjectives perçues entre deux fréquences ou deux timbres vocaliques, et du fait que de nombreux détails spectraux observables sur les spectrogrammes n'étaient pas pertinents d'un point de vue perceptif, les chercheurs ont eu recours à l'utilisation d'échelles psycho-acoustiques de fréquence (exprimées en Mels ou en Barks) (Zwicker, 1982) pour une représentation spectrale des sons de la parole (*cf.* figure 15).

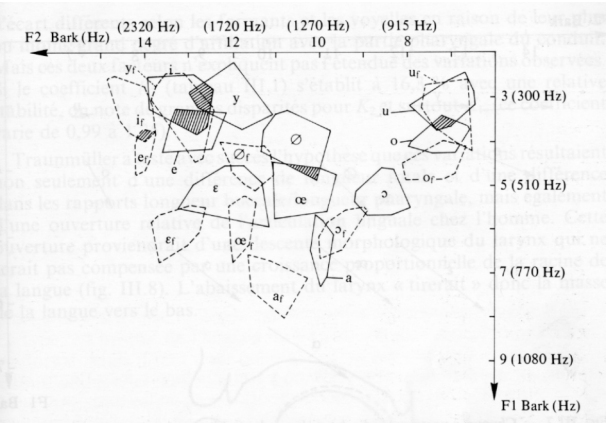


Figure 15
Zones de dispersion des voyelles orales du français sur le plan F1/F2 (échelle de Bark) extrait de Calliope (1989)

En ce qui concerne les mesures acoustiques, et malgré les réserves émises par Lindblöm (1962), il est incontestable que l'avènement du *'Sonagraphe'* allait grandement faciliter la tâche du chercheur. En effet, ce dernier pouvait mesurer la fréquence des formants, soit en prenant pour référence le centre de la barre horizontale dans le cas d'une analyse en bande large à 300 Hz, soit en sélectionnant un harmonique 'amplifié' en bande étroite à 45 Hz. De plus, et afin de disposer d'une représentation plus fine de la structure harmonique, il pouvait faire une section d'amplitude en un point précis du segment vocalique ou du segment consonantique (figure 16).

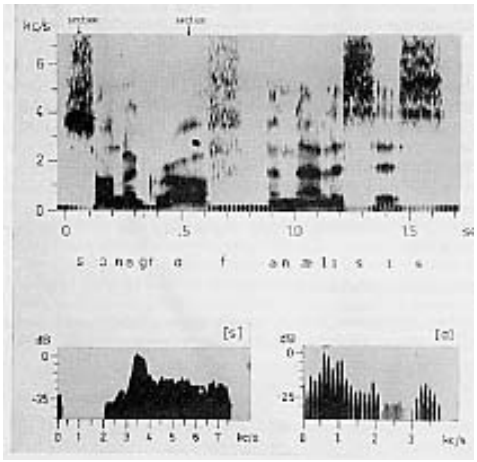


Figure 16
Spectrogrammes et sections d'amplitude effectués sur la consonne [s] et sur la voyelle [a], extrait de Fant (1968)

Les facilités offertes par la méthode spectrographique pour la visualisation et l'analyse du signal de parole ont eu pour résultat que les études d'acoustique ont pris une dimension nouvelle. Alors que jusqu'aux années 1950, celles-ci avaient été le plus souvent limitées à l'analyse des réalisations d'un nombre restreint de locuteurs, la spectrographie allait permettre d'étendre l'analyse acoustique à une population beaucoup plus étendue. De ce point de vue, l'étude pilote de Peterson & Barney (1952), qui associe analyse acoustique et traitement statistique, allait servir de modèle à de nombreuses études sur la variabilité inter et intra-locuteurs. Et en fait, depuis l'étude pilote de M. Joos (1948), premier linguiste à avoir utilisé la spectrographie jusqu'aux études les plus récentes (Maddieson & Ladefoged, 1996), il n'existe, à notre connaissance, peu ou pas d'études descriptives des voyelles qui n'aient été réalisées sans être fondées sur la méthode spectrographique.

2.1.2. Les consonnes

Les informations recueillies d'après l'examen des tracés kymographiques avaient permis de progresser dans la connaissance de ce type de sons. En effet, Rousselot et ses disciples avaient remarqué que la ligne de pression buccale était caractérisée par un décrochage de ladite ligne au moment de la phase d'explosion des consonnes occlusives, alors que la ligne laryngée était caractérisée par la présence d'une ondulation aléatoire de faible amplitude sur la tenue des consonnes fricatives, sans qu'il soit toutefois possible de déterminer la fréquence ou le niveau d'intensité de ces bruits. En ce qui concerne la durée, l'analyse avait été plus fructueuse du fait que la mesure des tenues consonantiques avait révélé des différences temporelles entre les consonnes sourdes et les consonnes sonores, différences que Rousselot *(op. cit.)* avait attribuées à un degré variable de force articulatoire.

Si utiles soient-elles, ces informations demeuraient limitées, car elles ne concernaient que les parties statiques des consonnes, et non pas les parties dynamiques, c'est-à-dire les transitions qui reflètent les déplacements articulatoires entre les consonnes et les voyelles. Et c'est précisément dans la représentation graphique de cette dimension dynamique de la parole que l'apport de la spectrographie allait se révéler déterminant. En effet, les chercheurs allaient rapidement s'apercevoir que les transitions des consonnes n'évoluaient pas sur l'axe du temps de façon aléatoire, mais au contraire de façon cohérente, et que la direction (positive, négative ou plate), ainsi que la pente (plus ou moins rapide) des transitions constituaient des indices acoustiques primaires du lieu et du mode d'articulation des consonnes (figure 17).

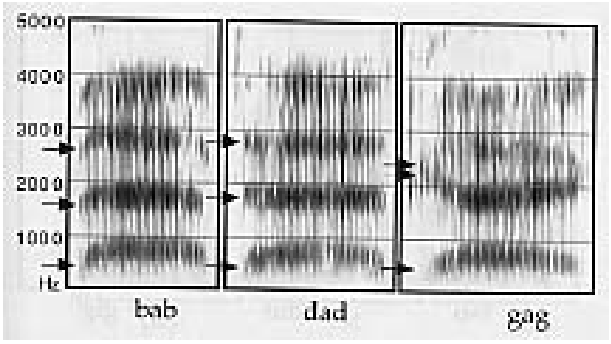


Figure 17
Spectrogrammes de consonnes occlusives [b, d, g] dans bab, dad, gag*, extrait de Ladefoged (2001)*

En fait, la plupart des indices acoustiques qui jouent un rôle dans la production et la perception des consonnes, allaient être découverts grâce à l'action des chercheurs des laboratoires Haskins aux États-Unis. Parmi ces indices, on citera en ce qui concerne :

- les consonnes occlusives : la durée des tenues, la direction et la pente des transitions, la fréquence des bruits d'explosion, la fréquence terminale des transitions.
- les consonnes fricatives : la durée des bruits, la direction et la pente des transitions, la fréquence des bruits.
- les consonnes nasales : la fréquence des formants de nasalité, la fréquence des anti-résonances ou zéros acoustiques, la réduction d'intensité des formants.
- les consonnes vocaliques : la durée des tenues et des transitions, la fréquence des formants et des transitions, la réduction d'intensité des formants, la continuité des formants, la présence ou l'absence de joints hauts ou bas etc.

On citera également les indices qui interviennent dans l'opposition consonne voisée-non voisée, c'est-à-dire la présence ou l'absence de la barre de voisement sur la tenue, la durée (relative) de la voyelle adjacente, l'intensité des bruits d'explosion ou des bruits de friction, la coupure (cutback) du premier formant, le délai d'établissement du voisement (Voice Onset Time), etc.

En résumé, on dira que la spectrographie a permis de recueillir entre 1950 et 1970 un nombre considérable d'informations à propos de la structure acoustique des sons du langage. Cette étape cruciale de la recherche qui a abouti à la découverte des principaux indices acoustiques, est liée au nom de P. Delattre dont les articles les plus marquants ainsi que ceux de ses collègues, ont fait l'objet d'une compilation dans les ouvrages de G. Fairbanks (1966) et d'I. Lehiste (1967).

2.2. La recherche appliquée à la synthèse et à la reconnaissance automatique

Universellement connue de tous les phonéticiens comme la méthode de référence pour la visualisation et l'analyse acoustique du signal de parole, la spectrographie n'était pas originellement destinée, tout au moins dans l'esprit de ses concepteurs, à être utilisée dans un but de recherche fondamentale. Lancé au début des années 1940, c'est-à-dire alors que la deuxième guerre mondiale venait d'éclater, le projet de visualisation de la parole avait une application militaire bien ciblée, celle de transmettre la parole non plus sous la forme d'un message oral qui aurait pu être intercepté et décodé, mais sous la forme d'un message visuel qui, en raison de son aspect novateur pour l'époque, aurait échappé à toute tentative de décodage. À la fin de la guerre, le projet n'avait pas été mené à terme, et de ce fait n'a pas trouvé son application militaire originelle. Plus tard, si la spectrographie a été utilisée par les services de la marine pour l'analyse de signaux sous-marins

permettant l'identification des navires de guerre sur la base des bruits émis par les moteurs ou les turbines, il n'en reste pas moins que cette utilisation de la spectrographie n'a été qu'occasionnelle dans cette application à vocation militaire. Par contre, la parole visualisée en termes de ses formes spectrales a été l'objet d'autres applications, notamment en synthèse et en reconnaissance de la parole.

2.2.1. L'application à la synthèse

C'est après que furent recueillies les informations concernant les indices acoustiques, que des ingénieurs travaillant en collaboration avec les linguistes, se sont intéressés à la fabrication d'appareils susceptibles de reproduire artificiellement la voix humaine. À ce sujet, on distinguera deux types de synthèse optique fondés sur le principe de la parole visualisée : la lecture des formes spectrales et la lecture de tracés paramétriques.

2.2.1.1. La synthèse par relecture de spectrogrammes

Si un appareil comme le *'Sonagraphe'* était capable de donner une représentation visuelle de la parole en termes de ses formes spectrales, il devait être possible d'effectuer l'opération inverse, en fabriquant une machine capable de restituer le signal vocal à partir de ces éléments d'informations. C'est le raisonnement qu'a tenu un ingénieur américain, F.S. Cooper, qui construisait en 1947 le premier synthétiseur de type optique, dont le principe reposait sur la conversion en ondes sonores des formes spectrales visibles sur un spectrogramme. La conception technique du 'Pattern Playback' était relativement simple. Une lumière émise par une lampe à arc traverse une roue tonale constituée de cinquante cercles concentriques d'opacité variable et ressort en autant de faisceaux lumineux modulés de 120 à 6000 Hz. Ces faisceaux sont concentrés par une lentille sur un miroir à 45° qui les renvoie vers une cellule photoélectrique. La lumière est soit transmise directement à travers le négatif d'une photographie originale, soit réfléchie par les traces de peinture blanche qui constituent une version simplifiée du spectrogramme. Dans l'esprit de son concepteur, cet appareil devait servir de machine à lire pour aveugles (Cooper, 1950). Toutefois, et malgré un taux d'intelligibilité relativement correct, il devint rapidement évident qu'il était difficile d'envisager la fabrication en série de ce type d'appareil, et encore moins de constituer une bibliothèque sous forme de photographies de spectrogrammes stylisés. C'est pourquoi le projet original d'aide aux aveugles par relecture de spectrogrammes a été rapidement abandonné. En revanche, ce type de synthèse allait devenir un outil de travail particulièrement efficace en recherche phonétique. En effet, la méthode fondée sur la peinture de traces blanches sur une bande de plastique, permettait de produire rapidement des stimuli synthétiques, de procéder à la modification instantanée de la fréquence et de la durée, et de juger le

résultat perceptif de cette modification manuelle. Sur un plan purement pragmatique, il est incontestable que le 'Pattern Playback' a été le complément idéal du *'Sonagraphe',* et le synthétiseur le plus utilisé pour la validation perceptuelle des indices acoustiques extraits à partir de l'analyse spectrographique (figure 18).

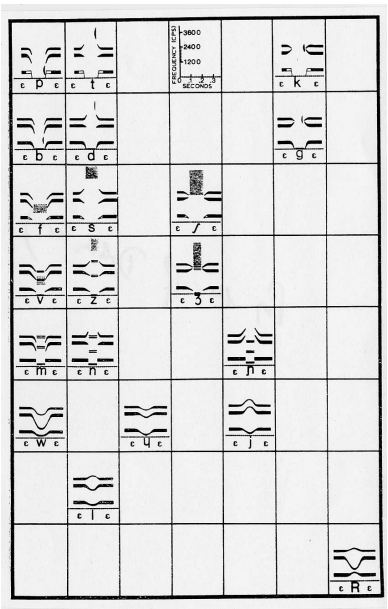


Figure 18
Tableau des formes spectrographiques stylisées des consonnes du français entre voyelles, extrait de Delattre (1970)

C'est ainsi que, après avoir démontré en un premier temps que les formants constituaient bien l'indice perceptif primaire responsable de la couleur vocalique (Delattre, 1951), les chercheurs du groupe Haskins allaient s'attacher à démontrer l'importance des transitions dans la perception des consonnes. Ces expériences dont l'objectif était la recherche d'invariants acoustiques, ont donné lieu à la formulation de la théorie du locus, selon laquelle les transitions convergent vers un point virtuel unique indépendamment du contexte vocalique, et constituent un indice majeur pour la perception du lieu d'articulation des consonnes (Delattre *et al.*, 1955).

Quoique cette méthode de synthèse fondée sur la recherche de l'invariance ait été critiquée pour avoir conduit à une hyper-simplification de la réalité acoustique brute, il n'en reste pas moins que les règles établies par Delattre et ses collègues (Delattre *et al.,* 1959) ont longtemps servi de référence en

matière de synthèse par règles (pour un bilan de la recherche effectuée durant les années 1950-1960, le lecteur pourra se reporter à la bibliographie commentée de Chafcouloff (1974)).

2.2.1.2. La synthèse à formants

Malgré ses avantages, dont le plus évident était sa facilité d'emploi, le 'Pattern Playback' présentait un certain nombre d'inconvénients. Il était impossible de procéder à des variations de la fréquence fondamentale, celle-ci étant fixée arbitrairement à 120 Hz, d'où l'impression déplaisante d'une voix monocorde ; de plus, le contrôle de l'intensité était peu précis, l'intensité étant proportionnelle à la quantité de peinture (blanche) réfléchie par la lumière. Enfin, la production des consonnes fricatives était déficiente en raison de l'absence d'une source de bruit.

La qualité auditive insuffisante de cette voix artificielle a incité les ingénieurs à se tourner vers un autre type de synthèse, en l'occurrence la synthèse à formants par laquelle la parole était reproduite non plus uniquement à partir de ses formes spectrales, mais à partir de tracés paramétriques concernant la fréquence fondamentale, les trois ou quatre premiers formants d'oralité, l'intensité globale, un ou deux formants de nasalité, la fréquence du bruit etc. En fait, le principe de la parole de synthèse de type optique restait le même que le précédent, si ce n'est que le nombre de paramètres était plus grand et le contrôle de chacun d'entre eux plus précis. Les tracés paramétriques étaient dessinés avec une encre conductrice d'électricité sur une feuille de plastique, et étaient lus par un lecteur photo-électrique ; leur conversion en ondes sonores était effectuée par l'intermédiaire de deux générateurs : un générateur de voisement pour les voyelles et les consonnes voisées, et un générateur de bruit pour les consonnes fricatives.

Plusieurs synthétiseurs à formants ont été construits entre 1960 et 1970, notamment au MIT (Massachusets Institute of Technology) de Boston, au RIT (Royal Institute of Technology) de Stockholm, *cf.* la série des différents modèles OVE (Orator Verbis Electricis) et le modèle PAT (Parametric Artificial Talker) construit à Edimbourg (voir la revue de Chafcouloff, *op.cit,* p. 106-148). En France, un synthétiseur de ce type a été construit par les ingénieurs de l'ENSERG à Grenoble (Paillé, Beauviala & Carré, 1970) et a été utilisé pour la synthèse de la première phrase de parole artificielle produite à l'Institut de Phonétique d'Aix-en-Provence, (Rossi & Chafcouloff, 1975), (*cf.* figure 19).

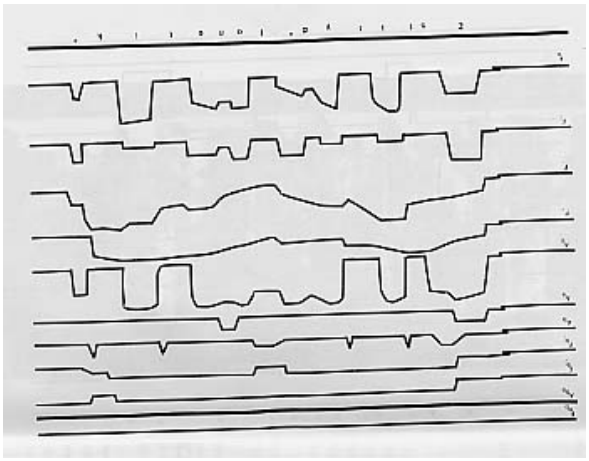


Figure 19
Évolution temporelle des tracés paramétriques de la phrase 'Institut de phonétique' sur synthétiseur à formants, extrait de Rossi et Chafcouloff (1975)

Avec l'avènement de l'informatique, ce type de synthèse 'artisanale' a cédé la place à une synthèse où la parole artificielle n'est plus produite à partir de tracés, mais à partir de données numériques correspondant aux paramètres, et qui s'affichent sur l'écran de l'ordinateur sous formes de tableaux. Comme l'opérateur dispose d'un nombre important de paramètres (trente-neuf exactement dans le logiciel de synthèse de Klatt (1980), et qu'il dispose de différents types de source vocale, la parole produite est d'une qualité auditive bien supérieure.

2.3. L'application à la reconnaissance de la parole

Il est de notoriété commune que la particularité du signal de parole est d'être foncièrement variable, et que cette variabilité intra ou inter-locuteurs constitue l'un des principaux obstacles rencontrés en reconnaissance automatique. Afin de surmonter cet obstacle, l'une des approches préconisées par certains chercheurs a été fondée sur l'utilisation des connaissances acoustiques, phonétiques et linguistiques acquises pendant les années 1950-1970, grâce à l'utilisation intensive du spectrographe et de son complément le relecteur de spectrogrammes. Cette approche peut être résumée comme suit. Si la machine était capable de reproduire artificiellement la parole par relecture de ses formes spectrales, il devait être possible de faire reconnaître ces mêmes formes par la machine, en les associant à des unités phonémiques spécifiques. En d'autres termes, il s'agissait de transférer et d'appliquer à l'ordinateur la compétence de l'expert phonéticien en reconnaissance

visuelle des formes (Cole & Zue, 1980). C'est à l'élaboration de systèmes experts, techniques couramment utilisées en intelligence artificielle, que se sont attachés les chercheurs adeptes de la reconnaissance analytique. Cependant, il s'est avéré que la performance humaine était de loin supérieure à celle de la plupart des systèmes de décodage acoustico-phonétique mis en œuvre durant la décennie 1970-1980. Les raisons de cette surperformance de l'homme sur la machine ont été exposées par Zue (1983) et tiennent aux faits suivants. L'expert émet un certain nombre d'hypothèses phonémiques émises sur la base de ses connaissances innées, et qui sont le fruit de son expérience et de sa culture linguistique. Dans ce but, il utilise des règles phonotactiques, allophoniques, phonologiques qui l'aident dans sa prise de décision. À la différence de la machine, la démarche de l'expert consiste à utiliser simultanément l'axe syntagmatique (temporel) et l'axe paragdimatique (combinatoire) pour effectuer une lecture soit « globale » ou « détaillée », ou encore une lecture « avant » ou « arrière » de l'image acoustique. Cette approche qui consistait à décoder l'information spectrographique par le biais d'une structure informatique était *a priori* séduisante, mais l'élaboration de tels systèmes de reconnaissance des formes s'est rapidement heurtée à deux obstacles majeurs.

Le premier obstacle concernait la quantité de données qu'il convenait de rassembler. En effet, le problème pour la machine est de nature quantitative au moment de la prise de décision. La constitution d'une base de connaissances, susceptible de résoudre les problèmes posés par la variabilité acoustique contextuelle, représente un travail considérable qui ne peut être accompli qu'au bout de (très) longues années de recherche fondamentale.

Le deuxième obstacle concerne la modélisation du raisonnement de l'expert. En effet, il convient de formaliser toutes ces connaissances sous formes de règles ou de méta-règles et, tâche encore plus complexe, de reproduire fidèlement sa démarche intellectuelle sous forme de séquences d'instructions à la machine.

En dépit de l'amélioration des connaissances au cours de ces dernières années, celles-ci sont restées trop qualitatives pour surmonter ces obstacles. Devant l'ampleur de la tâche à accomplir et la complexité des problèmes à résoudre, les systèmes de reconnaissance analytique ont rapidement cédé la place à des systèmes de reconnaissance globale fondés sur la reconnaissance de mots ou de vocabulaires de plus en plus étendus grâce à l'augmentation de la capacité de mémoire des ordinateurs. De plus, l'intérêt clairement affiché de certaines sociétés comme IBM, Texas Instruments ou Hewlett Packard de proposer des systèmes de reconnaissance vocale 'grand public' à faible coût, a rendu obsolète la réalisation de systèmes basés sur l'application de règles formelles de reconnaissance des formes acoustiques. La commercialisation de systèmes de reconnaissance fondés sur la modélisation statistique de la parole par Modèles de Markov cachés (HMM) ou par

réseaux de neurones artificiels (ANN) montre que dans ce domaine, la reconnaissance des formes acoustiques n'est plus ou peu d'actualité, et que d'autres solutions plus rentables à court terme ont été trouvées, par exemple le système Via Voice et autres systèmes de reconnaissance vocale actuellement disponibles sur le marché.

2.4. L'application à la thérapie de la parole

Le concept original d'une parole transcripte sous la forme de symboles susceptibles d' être déchiffrés et interprétés non pas par le biais de l'oreille, mais par celui de l'œil n'est pas nouveau, loin s'en faut et date de plus d'un siècle. En effet, c'est en 1867 que Melville Bell (dont le fils Graham Bell fût l'inventeur du téléphone), présentait pour la première fois un alphabet où chaque symbole manuscrit était associé à un son particulier du langage (figure 20).

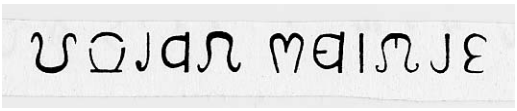


Figure 20
*'Les mots '*Visible Speech' *selon les symboles manuscrits employés par Melville Bell en 1867, extrait de Potter* et al. *(1947)*

À cette époque, la présentation de cet alphabet avait été saluée comme une étape importante vers la réalisation de ce qui avait vocation à devenir un système universel de communication entre les hommes, mais aussi comme un pas décisif vers la rééducation de personnes atteintes de surdité profonde ou partielle. Cependant, les expériences d'apprentissage montrèrent rapidement que ces formes symboliques n'étaient interprétrables par l'œil que sous certaines conditions, et que seuls certains monosyllabes ou bisyllabes, affranchis de tout contexte, pouvaient être mémorisés par les patients. À la suite de cet échec, et après que la kymographie et l'oscillographie eûssent montré leurs limites dans l'identification visuelle des sons du langage, la visualisation de la parole par la spectrographie semblait constituer une alternative dans cette optique thérapeutique. Dans ce but, les ingénieurs de la 'Bell Telephone Company' allaient conçevoir un système de 'traduction directe' où la parole visualisée était projetée en continu sur un écran de plastique au phosphore placé devant les patients, ceux-ci s'efforçant de reconstruire le message en associant les dites formes spectrales aux sons de leur système linguistique (figure 21).

Figure 21
Modèle d'un traducteur de parole visualisée, extrait de Potter et al. *(1947)*

Le concept paraissait prometteur, mais là encore, les limites de cette méthode n'allaient pas tarder à apparaître. L'identification des mots nécessitait un temps d'apprentissage long et le débit de parole devait être ralenti afin que les formes acoustiques puissent être interprétées par l'œil. De plus, l'information visuelle ne pouvait être utilisée par le patient pour effectuer simultanément la correction de sa propre production phonique. En effet, celui-ci ne parvenait à produire qu'*a posteriori* une forme acoustique plus ou moins semblable à la forme originale projetée sur l'écran. Enfin, comme nous l'avons mentionné à propos de la tentative de M. Bell, le vocabulaire était limité à des mono ou bi-syllabes (pour un exposé plus complet de la spectrographie appliquée à la rééducation des mal-entendants, on consultera l'article de Cole *et al.* (1980).

Cependant, si la visualisation des variations spectrales avait montré ses limites dans cette application à la rééducation des malentendants, elle n'en conservait pas moins tout son attrait en tant que méthode pour l'analyse acoustique des signaux de parole des patients souffrant d'un dysfonctionnement physio-pathologique du mécanisme de production. C'est effectivement dans cette fonction d'analyse qu'elle allait conduire au recueil d'informations précieuses concernant l'aide au diagnostic et le traitement de différents troubles de la parole.

2.4.1. Les dysarthries

L'étude pilote de Lehiste (1965) allait véritablement marquer le départ de la recherche dans le domaine de la thérapie de la parole. En effet, ce travail dans lequel l'auteur utilisait (de façon bizarre) des traits articulatoires pour tenter d'expliquer la production incorrecte de dix patients dysarthriques, allait être suivi de nombreux travaux où l'approche acoustique allait être privilégiée (*cf.* la synthèse des travaux faite par Ball et Code (1997)). Les procédures étaient fondées soit sur l'analyse d'un paramètre ou d'un indice déterminé pour un type de dysarthrie, lui-même défini à l'avance, soit sur la comparaison de paramètres ou d'indices entre différents types de dysarthries. Parmi les paramètres et indices qui ont retenu l'attention des chercheurs, on retiendra sur le plan segmental :

- la durée des tenues vocaliques et consonantiques dans différentes conditions de mot et d'accent, la durée des transitions de formants, la durée du VOT ;
- l'évolution des trajectoires de formants, la fréquence de départ et d'arrivée des transitions, le bruit d'explosion des consonnes occlusives ;
- l'intensité mesurée sur le mot ou la phrase, le rapport d'amplitude entre les pics spectraux, le rapport d'amplitude de la composante orale par rapport à la composante nasale etc.

Sur le plan suprasegmental, l'observation de la courbe mélodique modélisée grâce à la méthode MOMEL (Hirst and Espesser, *op. cit.)* a été porteuse d'enseignements pour l'établissement d'une dysprosodie chez le patient parkinsonien ou ataxique. Pour le premier, on a observé une diminution de la dynamique de F0, alors que pour le second, on a constaté une augmentation de la dynamique avec un accroissement des points-cibles par rapport à un sujet normal (Teston, 2000a).

2.4.2. Les apraxies ou les aphasies

En ce qui concerne les segments phoniques produits par des patients souffrant de ces pathologies, ce sont les mêmes indices qui ont été pris en compte. De plus, les chercheurs ont porté une attention particulière à la détérioration de l'échelle des fréquences et ont mesuré les retards de coarticulation (surtout la coarticulation de type anticipatoire), le nombre de syllabes produites par seconde, la durée des pauses et des silences, etc.

2.4.3. Le bégaiement

Pendant les années 1970, des études ont été entreprises sur des sujets affectés de bégaiement léger ou profond. Les mesures acoustiques ont porté sur la durée des segments voisés et bruités, ainsi que sur les variations temporelles dues à l'accélération ou au contraire au ralentissement du débit. Par ailleurs, on a analysé le timbre des voix sur la base des fréquences de formants, et on a examiné

dans quelle mesure les voyelles étaient plus ou moins centralisées. Toutefois, si la recherche acoustique a été active, on déplore quand même un certain manque de cohérence entre les résultats obtenus. En effet, les résultats manquent de cohérence et sont quelquefois peu comparables, du fait qu'il existe de nombreux facteurs de variabilité comme l'âge des patients, le type de bégaiement, la durée du traitement thérapeutique, le type de stimuli utilisés dans les tests, ainsi que la procédure analytique.

2.4.4. La surdité

La plupart des études ont porté sur l'analyse des voyelles afin de délimiter la zone de dispersion des formants. Sur le plan suprasegmental, la hauteur moyenne de F0 ainsi que les contours intonatifs ont été pris en compte. Enfin, et avec la banalisation 'relative' de l'implantation cochléaire chez les sourds profonds, les thérapeutes ont procédé ces dernières années à un suivi de l'état de l'acuité auditive du patient sur la base de mesures acoustiques.
En ce qui concerne les systèmes de visualisation de la parole dans une optique de rééducation des mal-entendants, le CRIN (Centre de Recherche en Informatique de Nancy) a mis au point un système de visualisation susceptible d'améliorer la phonation chez l'enfant sourd. Avec l'aide d'un orthophoniste, celui-ci apprend à maîtriser la prononciation des segments phoniques de sa langue, d'après les images des sons projetées sur l'écran d'un oscilloscope ou d'un téléviseur (figure 22).

Figure 22
Visualisation de l'image des sons émis par un enfant mal-entendant sous le contrôle d'un orthophoniste, extrait de Ferretti et Cinare (1984)

2.4.5. L'hypernasalité

Au même titre que la nasalité 'normale' dont la recherche d'invariants s'était soldée par un échec (Curtis, 1968), les résultats décevants obtenus par Bloomer & Peterson (1956), ainsi que par Dickson (1962) à propos de l'hypernasalité ont montré que la méthode spectrographique se prêtait

mal à l'analyse acoustique d'un phénomène qui demeure avant tout 'perceptuel'. Devant ce constat d'échec, Stevens *et al.* (1975) ont mis au point une méthode de visualisation de l'hypernasalité chez des enfants sourds. La méthode est fondée sur l'emploi d'un accéléromètre miniaturisé fixé sur les ailes du nez du patient, lequel permet de capter les vibrations de la muqueuse en réponse au passage de l'air dans le conduit nasal. Le signal de sortie s'inscrit en temps réel sur un oscilloscope ou sur un écran d'ordinateur. Le système permet l'affichage d'une courbe référence dite de normalité, que le sujet essaye de reproduire le mieux possible. Des mesures quantitatives du niveau de nasalisation ont été obtenues d'après l'analyse de mots contenant des consonnes nasales et qui sont prononcés soit à l'état isolé, soit en contexte de phrase (figure 23).

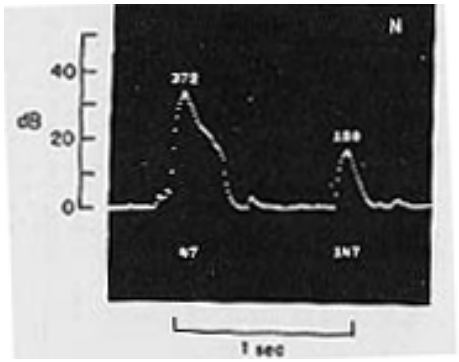


Figure 23
Affichage de la courbe de nasalité dans des mots avec (à gauche) et sans (à droite) consonnes nasales, extrait de Stevens et al. *(1976)*

2.4.6. La raucité

C'est vers les années 1970 que les questions posées par la qualité auditive de la voix humaine ont commencé à être traitées. Qu'il s'agisse d'une voix rauque ou éraillée dans les cas les plus bénins, d'une extinction de la voix due à une paralysie partielle ou totale des cordes vocales ou encore de la voix œsophagienne suite à l'ablation des cordes vocales dans les cas les plus graves, nombreuses ont été les études menées dans ce domaine (*cf.* la revue détaillée de Baken, 1987).
Là encore, les méthodes d'évaluation de la qualité vocale ont été essentiellement fondées sur la méthode spectrographique : l'analyse en bande étroite de voyelles tenues (200ms) ou encore le calcul moyen des spectres échantillonnés pendant un laps de temps de 1 à 3 secondes. Plusieurs types de raucité ont été mis en évidence : raucité légère selon que les formants des voyelles étaient plus ou moins mélangés à du bruit, raucité sévère selon que les impulsions glottales étaient

masqués par le bruit, ce qui sous-entend une prédominance de la source de bruit par rapport à la source vocale. Toutefois, comme pour le bégaiement, le manque d'homogénéité des résultats (en particulier à cause des différences de F0 entre les patients) a incité les thérapeutes à utiliser d'autres méthodes ; parmi celles-ci, on citera la méthode qui consiste à établir un rapport harmonique/bruit (H/N) exprimé en dB où l'amplitude moyenne de l'onde glottale est divisée par l'amplitude des composantes du bruit (Yumoto *et al.* 1984).

En résumé, on peut dire que la recherche appliquée à la pathologie du langage, au même titre que la recherche fondamentale sur la parole dite 'normale', a largement profité des méthodes de visualisation de la parole. Pour l'essentiel, les recherches menées aussi bien en milieu universitaire qu'hospitalier, ont porté :

- sur la description des caractéristiques acoustiques d'un trouble spécifique de la voix ;
- sur l'adoption de nouvelles techniques pour la mesure des paramètres physiques ;
- sur la comparaison des caractéristiques vocales avant et après traitement.

Cependant, il est incontestable que le recueil de données fondées sur la prise en compte des paramètres acoustiques ne présente pas toutes les garanties de fiabilité. C'est pourquoi les thérapeutes ont eu recours à d'autres types de paramètres comme les paramètres aérodynamiques associés à la respiration et à la phonation, et qui sont analysables en termes de données quantitatives. Parmi ces paramètres, le débit d'air buccal qui permet de mesurer la fuite glottique, laquelle influe sur le rendement laryngien, la pression intraorale (PIO) qui permet d'évaluer le forçage vocal (Teston, 2000b) et le débit d'air nasal qui permet d'évaluer l'hypernasalité sont les paramètres les plus couramment utilisés. Enfin, l'accent a été mis sur la nécessité de compléter l'analyse acoustique, aérodynamique ou physiologique, par un examen visuel fondé sur des techniques vidéo modernes comme la laryngoscopie, l'endoscopie ou la stroboscopie.

Conclusion

Au terme de ce travail consacré aux différents procédés et méthodes élaborés par les chercheurs et hommes de science pour donner une représentation visuelle de la parole, il apparaît comme une évidence que le long cheminement vers la connaissance du phénomène 'parole' s'est fait à un rythme variable en fonction du temps. À la période initiale d'acquisition des notions physiologiques de base sur le fonctionnement de l'appareil phonatoire redevables aux grammairiens et aux philosophes de l'antiquité, a succédé une longue période de stagnation et d'obscurantisme scientifique qui s'étend en fait du Moyen Age jusqu'à la deuxième moitié du XIXe siècle. En effet, c'est seulement à partir des années 1850-1900 que s'est produit l'évènement majeur

qui a vu l'introduction de la méthode graphique grâce au kymographe, lequel a permis d'établir la réalité physique de la parole, en tant que substance sonore analysable et décomposable en paramètres acoustiques, articulatoires ou aérodynamiques. À partir du moment où la parole n'était plus un phénomène abstrait, mais devenait au contraire un phénomène concret, la connaissance des faits acoustiques allait s'accélérer rapidement avec la construction d'appareils fonctionnant à l'électricité, et en particulier avec l'oscillographe couplé à un détecteur de fréquence fondamentale ou d'intensité pour la mesure des variations de ces paramètres. Dès la fin de la deuxième guerre mondiale, l'introduction de la méthode spectrographique allait entraîner une accélération de l'effort de recherche, lequel allait encore s'accentuer avec l'avènement de l'informatique et l'adoption généralisée des méthodes de traitement numérique d'analyse et de synthèse du signal de parole.

Parmi les méthodes de visualisation que nous avons décrites, et sans sous-estimer nullement l'apport de la kymographie et de l'oscillographie, il est incontestable que c'est l'introduction de la spectrographie qui a constitué le moment clé dans l'histoire de la recherche phonétique, dans la mesure où elle a permis aux chercheurs de combler en quelques années le retard de connaissances qu'ils avaient accumulé au cours des siècles précédents dans le domaine de l'acoustique par rapport au domaine physiologique.

Du fait qu'elle présentait l'avantage de donner une représentation graphique tri-dimensionnelle de la parole, la spectrographie a permis l'analyse des propriétés spectrales des sons, et a été à l'origine de la découverte des principaux indices acoustiques porteurs d'information pour l'identification phonémique. Dotés d'un outil de travail performant, les chercheurs se sont attelés dès lors au travail de longue haleine que représentait la description phonétique des systèmes phonologiques des langues en se fondant sur des critères acoustiques. De plus, comme la spectrographie a été souvent utilisée en synchronisation avec des techniques d'investigation physiologique comme la radio-cinématographie ou l'aérométrie, elle a permis d'établir des corrélations entre faits acoustiques et faits articulatoires ou aérodynamiques, qui se sont avérées fort utiles sur le plan pédagogique, et en particulier dans la didactique des langues étrangères.

Sur le plan linguistique, le recueil des données acoustiques, si difficile soit-il pour extraire de ces formes spectrales 'visibles' les informations 'pertinentes' sur le plan perceptif, a été à l'origine de l'adoption d'un nouveau type de classification phonétique. Alors que les sons du langage avaient été classés jusqu'alors selon des critères articulatoires, Jakobson *et al.* (1952) ont préconisé l'adoption d'un système original de classification des phonèmes (par opposition binaire) fondé sur des traits acoustiques 'distinctifs', classement dont le fondement repose sur une analyse fine des propriétés spectrales, et en particulier de la distribution de l'énergie acoustique. Quoique cette classification soit souvent demeurée abstraite, du fait qu'il existe une rupture de correspondance

biunivoque entre les plans acoustique et perceptif, laquelle a été dénoncée par Rossi (1977), il n'en demeure pas moins qu'elle constitue une étape importante dans le choix des critères retenus pour l'inventaire et le classement des sons des langues du monde.

Comme ces traits acoustiques constituaient une réalité phonétique robuste et que nombre d'entre eux se retrouvaient dans des langues d'appartenance et d'origine géographique diverses, les chercheurs allaient parallèlement s'attacher à démontrer l'existence d'universaux linguistiques. En ce sens, les études de Lindblöm (1963) et de Delattre (1965) sur le phénomène de centralisation des voyelles selon le débit et de l'accent, ainsi que celles d'Öhman (1966) sur l'étendue des faits de coarticulation entre les gestes vocaliques et consonantiques dans des séquences VCV ont fait date, et les données acoustiques recueillies par ces auteurs ont servi de fondement à l'élaboration de modèles de production de la parole.

Cependant, malgré une utilisation intensive de la méthode spectrographique, celle-ci n'a pas permis, loin s'en faut, d'apporter de réponse claire et non ambiguë à de nombreuses questions que se posaient les chercheurs. Comme l'ont signalé certains d'entre eux, notamment Lindblöm (1962), l'un des principaux inconvénients de la spectrographie résidait dans la trop grande richesse d'informations soumises à l'œil du chercheur. Si certaines d'entre elles ont retenu son attention, il n'en reste pas moins d'autres qui, soumises à des modifications importantes en fonction des influences segmentales ou suprasegmentales, ont échappé à son observation, si fine soit-elle. En réalité, si l'on se fonde sur la théorie quantique de Stevens (1972), selon laquelle un déplacement articulatoire de forte amplitude n'entraîne pas nécessairement une variation acoustique d'ampleur équivalente, il semble acquis que certaines variations du signal n'apparaissent pas dans sa représentation spectrale, et doivent être assimilées sur le plan articulatoire à des 'gestes cachés', confirmant de ce fait le manque de correspondance acoustico-articulatoire dans l'organisation gesturale dynamique de la parole (Browman & Goldstein,1990).

Dans ces conditions, il est apparu comme une évidence que la méthode spectrographique, si performante soit-elle, renfermait ses propres limites et ne constituait pas la panacée universelle aux problèmes que rencontraient les chercheurs dans le domaine de l'investigation acoustique. D'autre part, il est permis de se poser la question de savoir si les chercheurs avaient tiré la quintessence des informations susceptibles d'être extraites d'une analyse spectrographique fine du signal de parole. Telle ne semblait pas être l'opinion de F.S. Cooper qui, à l'occasion d'un voyage d'études que nous effectuions aux États-Unis, nous confiait que, à son opinion, la recherche sur les indices acoustiques de la parole, avait été interrompue de façon prématurée, et que beaucoup restait à découvrir dans ce domaine. Cette remarque faite au cours de l'été 1975, se justifiait par le fait que, après la mort de P. Delattre, peu de chercheurs se sont donné pour tâche de poursuivre l'œuvre des pionniers du groupe

Haskins. Pourtant il était apparu que leur méthode d'analyse des patrons spectraux et de vérification par synthèse était critiquable en ce qu'elle pouvait donner lieu à la création de véritables 'monstres acoustiques' peu en rapport avec la réalité acoustique brute. En règle générale, peu d'études ont été entreprises pour remettre en cause la validité de certains de ces indices ; parmi celles-ci, on citera celles de Fischer-Jorgensen (1954) à propos des consonnes occlusives du danois et de Chafcouloff (1983) à propos de l'utilisation d'indices naturels ou artificiels (en d'autres termes de vrais ou faux indices) pour la synthèse des consonnes vocaliques du français.

En deuxième lieu, la remarque de F.S Cooper s'explique par le fait que, après des années de travaux intensifs, les efforts des chercheurs n'avaient pas été toujours couronnés de succès, en particulier dans le domaine de l'identification des voix individuelles. En effet, si la méthode spectrographique s'était révélée fort utile pour l'extraction d'indices et la production par synthèse d'une parole intelligible à partir de ces données, l'analyse allait s'avérer insuffisante dans le cadre d'une application à la reconnaissance et à l'identification de l'individu. En effet, s'était posée à cette époque la question des empreintes vocales (sous la forme de patrons spectraux et autres informations obtenues grâce à l'analyse spectrographique) et de leur utilisation éventuelle devant les tribunaux au même titre que les empreintes digitales. Subventionnées par le ministère de la Justice, des études allaient être entreprises à des fins d'identification juridique en particulier aux États-Unis. Toutefois, ces recherches n'allaient aboutir qu'à des résultats controversés, et plusieurs chercheurs de renommée mondiale aux États-Unis (Bolt *et al.,* 1970) ainsi que plusieurs membres du GFCP (Groupe Francophone de la Communication Parlée) en France (Boë *et al.,* 1999) allaient prendre une position très ferme contre la validité de ce type d'informations recueillies par des 'experts scientifiques' sur la base d'une analyse des empreintes vocales. Le fait que la fiabilité de ces expertises ait été mise en doute, ajouté aux découvertes faites ces dernières années dans le domaine de la structure génétique de l'ADN chez l'individu (Jeffreys *et al.,* 1985), a eu pour résultat que ce type de recherche sur les caractéristiques individuelles de la voix des locuteurs a perdu tout intérêt du point de vue de son application juridique ou médico-légale, et a été de ce fait abandonné.

Il n'en demeure pas moins que si la recherche n'a pas été poursuivie dans cette optique d'application bien précise, elle est restée fort active dans bien d'autres domaines, et en particulier en synthèse et en reconnaissance automatique de la parole.

En synthèse à partir du texte, une collaboration scientifique étroite entre les chercheurs de diverses disciplines (ingénierie, informatique, acoustique, statistique, linguistique, psychologie) a donné lieu à la publication de nombreux travaux qui portent notamment sur la modélisation de différentes sources glottiques, la conversion graphème/phonème, c'est-à-dire la conversion du texte écrit en une représentation linguistique appropriée, la stylisation des contours mélodiques, l'évaluation

auditive de la voix artificielle etc. (Van Santen *et al.,* 1997). Ces travaux sont importants en particulier pour la synthèse par concaténation, dont la voix, si elle est intelligible, manque par contre de naturel, en raison d'une transplantation prosodique souvent inadaptée. Dans ce domaine, des études complémentaires sont entreprises pour adopter les patrons intonatifs et rythmiques adéquats afin d'atteindre un meilleur rapport intelligibilité/naturel de la voix. En outre, il est certainement possible d'affiner la qualité segmentale par une meilleure modélisation des passages entre parties voisées et non voisées.

En reconnaissance automatique, l'incapacité actuelle des modèles existants d'extraire et de modéliser la variabilité du signal acoustique constitue assurément un obstacle majeur. C'est pourquoi les systèmes les plus performants en la matière sont des systèmes hybrides de type probabiliste HMM/ANN fondés sur le traitement de données statistiques et la reconnaissance d'unités lexicales et de vocabulaires de plus en plus étendus grâce à l'augmentation de capacité de mémoire des ordinateurs. Néanmoins, ces systèmes renferment leurs propres limites, constatation qui a poussé certains chercheurs à proposer des solutions alternatives pour résoudre le problème de la reconnaissance robuste. C'est notamment le cas de Bourlard (1996) qui remet en cause l'approche dominante par HMM et préconise une approche fondée sur la notion d'accepteur stochastique à nombre d'états fini (SFSA) ainsi qu'une approche en sous-bandes de fréquences susceptible de conduire à l'obtention des taux de reconnaissance intéressants. Néanmoins et de l'aveu même de ce dernier auteur, il est indéniable que dans le domaine de la reconnaissance automatique, beaucoup de temps sera nécessaire pour déboucher un jour sur des solutions (quasi) optimales. Cette réflexion peut s'appliquer pareillement à d'autres domaines de la recherche phonétique, laquelle a progressé à pas de géant et a permis d'accéder à une meilleure connaissance de la production, de la perception et de la transmission de la voix humaine grâce aux différentes méthodes de visualisation de la parole conçues et réalisées au cours des siècles précédents. Nul doute que de nouvelles méthodes seront mises au point dans un avenir proche, et qui contribueront certainement à affiner nos connaissances. De nombreux obstacles restent à surmonter, et d'autres défis stimulants attendent le chercheur, mais en l'état, il convient de faire preuve d'humilité et de reconnaître que le phénomène 'parole' est loin d'avoir livré tous ses secrets.

Remerciements
à Michel Pitermann pour la relecture du texte et ses remarques pertinentes.

Références bibliographiques

BAKEN, R.J. (1987), *Clinical measurement of Speech and Voice*, Taylor and Francis Ltd, London, 518 p.

BLOOMER, H., PETERSON, G. (1956), A Spectrographic Study of Hypernasality, *Cleft Palate Bulletin*, 6 (2), p. 10-12.

BOË, J.-L., BIMBOT, F., BONASTRE, J.-F. et DUPONT, P. (1999), De l'évaluation des systèmes de vérification du locuteur à la mise en cause des expertises vocales en identification juridique, *Langues,* vol. 2 (4), p. 270-288.

BOLT, R., COOPER, F.S., DAVID, E.E., DENES, P. B., PICKETT, J. M. & STEVENS, K. N. (1970), Speaker Identification by Speech Spectrograms : a scientists' view of its reliability for legal purposes, *Journal of the Acoustical Society of America*, 47 (2) II, p. 597-612.

BOURLARD, H. (1996), Reconnaissance automatique de la parole : modélisation ou description, *XXIèmes Journées d'Etudes sur la Parole*, Avignon, Centre d'Enseignement et de Recherche en Informatique, p. 263-272.

BROWMAN, C.P., GOLDSTEIN, L. (1990), Representation and reality : physical Systems and phonological Structure, *Journal of Phonetics,* 18, p. 411-424.

BRÜCKE, J. (1856), *Grundzüge der Physiologie und Systematik der Sprachlaute für Linguisten und Taubstummenlehrer*, Auflage, Wien.

CALLIOPE (1989), (éd.), *La parole et son traitement automatique*, Masson, 717 p.

CHAFCOULOFF, M. (1974), *Vingt-cinq années de recherche en synthèse de la parole*, éditions du CNRS, 287 p.

CHAFCOULOFF, M. (1983), Indices naturels et indices atificiels en parole de synthèse, *Phonetica*, 40, p. 293-310.

COLE, R.A, RUDNICKY, A.I., ZUE, V.W., REDDY, D.R. (1980), Speech as Patterns on Paper, in Cole, R.A., ed., *Perception and Production of fluent Speech*, Lawrence Erlbaum, Hillsdale, N.J., chapter 1, p. 3-50.

COLE, R.A, ZUE, V.W. (1980), Speech as eyes see it, in Nickerson, R.S., ed., *Attention and Performance VIII*, Lawrence Erlbaum, Hillsdale, N.J., p. 475-494.

COOPER, F.S. (1950), Research on Reading-machines for the Blind, in *Blindness: modern approaches to the unseen environment*, Princeton University Press, p. 512-543.

CRANDALL, I.B. (1925), Sounds of Speech, *Bell System Technical Journal*, 4, p. 586-626.

CURTIS, J.-P. (1968), Acoustics of Speech Production and Nasalization, in D.C. Spriesterbach, D. Sherman, eds, *Cleft Palate and Communication*, 27-60, Academic Press, New-York.

DELATTRE, P.C. (1951), The use of the Pattern Playback in Studies of vowel color by Synthesis, *Journal of the Acoustical Society of America,* vol. 22 (5), p. 678.

DELATTRE, P.C (1970), Des indices acoustiques aux traits pertinents, *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague 6-13th september 1967, p. 35-47, B. Hala *et al.,* eds, Academia Publishing House of the Czechoslovak Academy.

DELATTRE, P.C., LIBERMAN, A.M., COOPER, F.S. (1955), Acoustic Loci and transitional Cues for Consonants, *Journal of the Acoustical Society of America,* vol. 27 (4) , p. 769-773.

DICKSON, D. R. (1962), An acoustic Study of Nasality, *Journal of Speech and Hearing Research,* vol. 5, p. 103-111.

DI CRISTO, A. (1975), *Soixante et dix ans de recherches en prosodie*, Etudes Phonétiques 1, Publications de l'Université de Provence.

DURAND, M. (1947), *Voyelles longues et voyelles brèves*, Collection Linguistique, 49, Klincsieck, Paris, 195 p.

FAIRBANKS, G. (1966), *Experimental Phonetics : Selected articles*, University of Illinois, Urbana, 274 p.

FANT, G. (1968), Analysis and Synthesis of Speech Process, in *Manual of Phonetics*, B. Malmberg, ed., North Holland Publishing Company, Amsterdam, Chapter 8, p. 173-277.

FARMER, A. (1997), Spectrography, in Ball, M.J., Code C., eds, *Instrumental Clinical Phonetics*, Chapter 2, Singular Publishing Group, Inc., San Diego, California, 296 p.

FERRETTI , M. et CINARE, F. (1984), *Synthèse, reconnaissance de la parole*, Edi Tests, 282 p.

FISCHER-JORGENSEN, E. (1954) 'Acoustic analysis of stop consonants', *Miscellanea Phonetica*, vol. II, 42-59.

FLETCHER, (1929), *Speech and Hearing*, New-York.

GEMELLI, A., PASTORI, G. (1934), *Acustica del linguaggio*, Milano.

GRAMMONT, M. (1933), *Traité de Phonétique*, Delagrave, Paris.

HELMHOLTZ, H. (1863), *Die Lehre von den Tonempfindungen,* Braunschweig.

HERMANN, L. (1895), *Weitere Untersuchsungen über die Wesen der Vokale*, Arch. für Physiologie, LXI 195, Pflüger.

HIRST, D. J., ESPESSER, R. (1993), Automatic modelling of fundamental frequency using a quadratic spline function, *Travaux de l'Institut de Phonétique d'Aix*, vol. 15, p. 75-85.

JAKOBSON. R., FANT, G., HALLE, M. (1952), *Preliminaries to Speech Analysis*, The MIT Press, Cambridge, Massachusetts.

JEFFREYS, A.J., WILSON, U. and THEIN, S.L. (1985), Individual specific fingerprints of human DNA, *Nature*, 316, 76-79.

JOOS, M. (1948), *Acoustic Phonetics*, Language Monograph, Linguistic Society of America, vol. 24, Baltimore, p. 1-136.

KLATT, D. H. (1980), Software for a cascade/parallel formant Synthesizer, *Journal of the Acoustical Society of America,* vol. 67, p. 971-995.

KOENIG, W., DUNN, H.K., LACY, L. (1946), The Sound Spectrograph, *Journal of the Acoustical Society of America,* 17, p. 19-49.

LADEFOGED, P. (2001), *Vowels and consonants : an introduction to the Sounds of Languages*, Blackwell Publishers, Oxford, 191 p.

LEHISTE, I. (1965), Some acoustic characteristics of disarthric speech, *Bibliotheca Phonetica*, 2, p. 1-124.

LEHISTE, I. (1967), *Readings in Acoustic Phonetics*, in I. Lehiste, ed., The MIT Press, Cambridge, Massachusetts.

LIBERMAN, A.M., INGEMAN F., LISKER, L., DELATTRE, P.C., COOPER, F.S. (1959), Minimal rules for Synthesizing Speech, *Journal of the Acoustical Society of America,* 31, p. 1490-1499.

LINDBLÖM, B. (1962), Accuracy and limitations of Sonagraph measurements, *Proceedings of the fourth International Congress of Phonetic Sciences*, Helsinki 1961, Mouton, The Hague, p. 188-202.

LINDBLÖM, B. (1963), Spectrographic study of vowel reduction, *Journal of the Acoustical Society of America,* 35, p. 1173-1781.

MADDIESON, I., LADEFOGED P. (1996)*, The sounds of the world's Languages,* Blackwell Publ., Oxford, U.K. 426 p.

MARTIN, P. (1986), Une méthode de calcul rapide du peigne spectral pour la mesure de la fréquence fondamentale, *Travaux de l'Institut de Phonétique d'Aix*, vol. 10, p. 359-369.

MARTIN, P. (1996), Winpitch : F0 en temps réel sous Windows, *XXIèmes Journées d'Etude sur la Parole*, (Avignon, 10-14 juin), Groupe Francophone de la Communication Parlée, p. 224-227.

MARTIN, P. (2005), Petite histoire de l'analyse de la fréquence fondamentale, *Un siècle de phonétique expérimentale : Histoire et développement. De Théodore Rosset à J. Ohala*, Colloque (24-25 février), Institut de la Communication parlée, Université Stendhal, Grenoble.

MERRY, G.N. (1921), Nasal resonance, *Quarterly Journal of Speech*, 7, p. 171-172.

ÖHMAN, S. (1966), Coarticulation in VCV utterances : spectrographic measurements, *Journal of the Acoustical Society of America,* 39, p. 151-168.

PAGET, Sir R. (1924), The nature and artificial production of consonant sounds, *Proceedings of the Royal Society of London*, A106, p. 150.

PAILLE, J., BEAUVIALA, J.-P., CARRE, R. (1970), Description et utilisation d'un synthétiseur du type à formants, *Revue de Physique appliquée*, 5, p. 785-793.

PANCONCELLI-CALZIA, G. (1957), Earlier History of Phonetics, in *Manual of Phonetics* , L. Kaiser, ed., North Holland Publishing Company, Amsterdam, p. 3-17.

PETERSON, G.E., BARNEY (1952), Control Methods used in a Study of Vowels, *Journal of the Acoustical Society of America,* 24, p. 175-184.

PIPPING, H. (1890), *Om klangfärgen hos sjungna vokaler,* Helsinki.

POTTER, R.K. KOPP, G.A., GREEN, H.G. (1947), *Visible Speech,* Dover Publications, New-York.

ROSSI, M. (1965), Contribution à l'étude des faits prosodiques dans un parler de l'Italie du Nord, *Langage et Comportement*, 1, p. 5-30.

ROSSI, M. (1971a), L'intensité spécifique des voyelles, *Phonetica*, 24, p. 129-161.

ROSSI, M. (1971b), Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole, *Phonetica*, 23, p. 1-33.

ROSSI, M. (1972), Le seuil différentiel de durée, *Papers in Linguistics and Phonetics to the memory of P. Delattre*, Mouton, The Hague, p. 436-450.

ROSSI, M. (1976a), *Contribution à la méthodologie de l'analyse linguistique avec application à la description phonétique et phonologique du parler de Rossano,* Thèse d'état, Librairie H. Champion, Paris.

ROSSI, M. (1976b), La perception des variations d'intensité*, Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 3, p. 361-457.

ROSSI, M. (1977), Les traits acoustiques, *La Linguistique*, 13, fasc. 1, p. 63-82.

ROSSI, M. (1978), The perception of non-repetitive intensity glides on vowels, *Journal of Phonetics*, 6, p. 9-18.

ROSSI, M., CHAFCOULOFF, M. (1975), La synthèse de la parole et la recherche dans le domaine de la phonétique expérimentale, *En hommage à G. Mounin*, CLOS, vol. 5, p. 345-364.

ROUSSELOT, l'abbé (1897), *Principes de phonétique expérimentale*, H. Welter, Paris.

SCRIPTURE, E.W. (1902), *The Elements of Experimental Phonetics,* AMS Press Inc.*,* Charles Scribner's sons, New-York.

SOVIJARVI, A. (1938), Die wechselnden und festen Formanten der Vokale erklärt durch Spektrogramme und Röntgengramme der finnischen Vokale, *Proceedings of the 3rd International Congress of Phonetik Sciences*, Gent, p. 407-420.

STEINBERG, J.-C. (1934), Application of Sound Measuring Instruments to the Study of Phonetic Problems, *Journal of the Acoustical Society of America,* 6, p. 16-24.

STEVENS, K. N (1972), The quantal nature of speech : Evidence from articulatory-acoustic data, in E.E. David, P. Denes, eds, *Human Communication; A unified view*, New-York, Mc Graw-Hill, p. 51-66.

STEVENS, K.N., KALIKOW, D.N., WILLEMAIN, T.R. (1975), A Miniature Accelerometer for Detecting Glottal Waveforms and Nasalisation, *Journal of Speech and Hearing Research,* 18, p. 594-599.

STEVENS, K. N., NICKERSON, R.S., BOOTHROYD, A. ROLLINS, A.M. (1976), Assessment of Nasalization or Nasality in the Speech of Children, *Journal of Speech and Hearing Research*, 19, p. 393-416.

SWEET, H. (1890), *Primer of Phonetics*, Oxford.

VAN SANTEN, J.-P., SPROAT, R.W., OLIVE, J.-P. and J. HIRSCHBERG, eds, (1997), *Progress in Speech Synthesis*, Springer, 598 p.

VIËTOR, W. (1923), *Elemente der Phonetik,* 4, Aufl., Leipzig.

TARNOCZY, T. (1948), Resonance data concerning Nasals, Laterals and Trills, *Word*, 4, p. 71-77.

TESTON, B. (1984), Un système de mesure des paramètres aéro-dynamiques de la parole : le Polyphonomètre, Modèle III, *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, vol. 9, p. 373-383.

TESTON, B., GALINDO, B. (1995), A diagnostic and Rehabilitation Aid Workstation for Speech and Voice Pathologies, *Proccedings Eurospeech,* 4, Madrid, European Speech Communication Association, p. 1883-1886.

TESTON, B. (2000a), L'évaluation objective des dysfonctionnements de la voix et de la parole, première partie : les dysarthries, *Travaux Interdisciplinaires du Laboratoire Parole et Langage,* vol. 19, p. 115-154.

TESTON, B. (2000b), Les dysfonctionnements pathologiques de la production de la parole, *La parole*, Hermès Science Publications, Chapitre 11, 409 p.

TRENDELENBURG, F. (1935), *Klänge und Geraüsche*, p. 138-150, Berlin und '*Einführung in die Akustik'*, Zweite Auflage, Berlin (1950), p. 359-362.

VIERORDT, K., LUDWIG, G. (1855), Beiträge zu der Lehre von den Atembewegungen, *Arch. Physiol. Heilkunde*, 14, p. 253-271.

YUMOTO, E., SASAKI, Y., OKAMURA, H. (1984), Harmonics-to-noise ratio and psychophysical measurement of the degree of Hoarseness, *Journal of Speech and Hearing Research,* 27, p. 2-6.

ZUE, V. (1983), The Use of phonetic Rules in Automatic Speech Recognition, *Speech Communication*, 2, p. 181-186.

ZWICKER, E. (1982), *Psychoakustik*, Springer, Berlin.